

Doble Grado Ingeniería Informática y Administración y Dirección
de Empresas
(2019 – 2020)

Trabajo Fin de Grado

“Herramienta para el análisis de difusión de noticias falsas en Twitter”

Ignacio Laviña Faustmann

Tutor:

Antonio Berlanga de Jesús



Esta obra se encuentra sujeta a la licencia Creative Commons **Reconocimiento – No Comercial – Sin Obra Derivada**

RESUMEN EJECUTIVO

Los bulos y las noticias falsas son un problema han existido desde tiempos antiguos, pero la era digital y la normalización del uso de las redes sociales han magnificado el alcance y los efectos del problema. Conceptos como la viralización o el anonimato en las redes sociales han potenciado el impacto de las noticias falsas, haciendo que agencias de noticias, empresas y gobiernos comiencen a tomar medidas para tratar de analizar o solucionar este problema.

Este proyecto trata de proporcionar una herramienta que ayude a todas aquellas personas o entidades que decidan afrontar este problema mediante el desarrollo de una aplicación web accesible para cualquiera que permita analizar la forma en la que se difunden las noticias en las redes sociales, centrando su funcionalidad en la red social Twitter. Para ello se ha construido una aplicación web que permite analizar la difusión de noticias falsas mediante herramientas de visualización de datos, además de la detección de cuentas bot, la agrupación automática de tweets relacionados con una misma noticia o el envío de respuestas a un número elevado de tweets de forma simultánea. La aplicación construida sirve como un mínimo producto viable que permite validar la viabilidad de la idea, de cara a hacer pruebas en un entorno real, recibir feedback y orientar las futuras versiones de la herramienta.

El presente documento trata de exponer el problema y su contexto, así como el proceso y la documentación técnica necesaria que pueda ayudar a reproducir dicho sistema, exponiendo al final los resultados de la aplicación construida, las posibles vías de mejora y oportunidades de explotación de la misma.

ÍNDICE

1.	RESUMEN Y ESTRUCTURA.....	1
1.1.	Resumen ejecutivo.....	1
1.2.	Keywords.....	1
1.3.	Objetivos y motivación	2
1.4.	Estructura.....	2
1.5.	Glosario	3
1.6.	Lista de Acrónimos.....	4
2.	INTRODUCCIÓN.....	6
2.1.	Introducción.....	6
2.2.	Relevancia de la tecnología en el problema de las fake news.....	6
2.3.	Propuesta de solución.....	8
2.4.	Descripción y alcance del sistema.....	8
2.5.	Capacidades generales.....	9
2.6.	Identificación de los interesados del sistema (stakeholders)	11
3.	CONTEXTO Y ESTADO DEL ARTE.....	14
3.1.	Valoración de la situación actual	14
3.2.	Contexto socio económico.....	15
3.3.	Marco regulador	16
3.4.	Estado del arte	19
4.	GESTIÓN DEL PROYECTO.....	25
4.1.	Metodología de desarrollo.....	25
4.2.	Planificación del proyecto.....	26
4.3.	Análisis de recursos del proyecto	30
4.4.	Análisis de costes del proyecto	33

5.	ANÁLISIS DEL SISTEMA.....	35
5.1.	Especificación de requisitos	35
5.2.	Especificación de casos de uso.....	49
5.3.	Casos de uso en formato expandido.....	51
5.4.	Matriz de trazabilidad de casos de uso.....	75
5.5.	Plan de gestión de la calidad.....	77
6.	DISEÑO DEL SISTEMA.....	78
6.1.	Arquitectura del sistema.....	78
6.2.	Entorno operacional	80
6.3.	Definición del modelo de datos	81
7.	DISEÑO DE LA INTERFAZ	89
7.1.	Prototipo de bajo nivel.....	89
7.2.	Panel de diseño o Mood Board.....	98
7.3.	Tecnologías utilizadas	100
8.	CONSTRUCCIÓN / RESULTADOS.....	103
8.1.	Resultados.....	103
8.2.	Futuras vías de exploración	116
8.3.	Limitaciones del proyecto	117
8.4.	Impacto socio económico	118
9.	CONCLUSIONES	119
10.	BIBLIOGRAFÍA.....	121
11.	ANEXO	125
11.1.	DECLARACIÓN DE ORIGINALIDAD	125
11.2.	DIAGRAMA DE GANTT.....	126
11.3.	DIAGRAMA DE CASOS DE USO	128
11.4.	MODELO DE INFORMACIÓN	129

ÍNDICE DE ILUSTRACIONES

Ilustración 1. Fase de planificación Diagrama de Gantt	27
Ilustración 2. Fase de planificación Diagrama de Gantt (Detalle ampliado).....	27
Ilustración 3. Fase de Desarrollo Diagrama de Gantt, Sprint 1 y 2	28
Ilustración 4. Fase de Desarrollo Diagrama de Gantt, Sprint 1 y 2 (Detalle ampliado)	28
Ilustración 5. Fase de Desarrollo Diagrama de Gantt, Sprint 3 y 4	29
Ilustración 6. Fase de Desarrollo Diagrama de Gantt, Sprint 3 y 4 (Detalle ampliado)	29
Ilustración 7. Fase de Desarrollo Diagrama de Gantt, Sprint 5 y 6	29
Ilustración 8. Fase de Desarrollo Diagrama de Gantt, Sprint 5 y 6 (Detalle ampliado)	30
Ilustración 9. Fase de Deployment Diagrama de Gantt	30
Ilustración 10. Diagrama de casos de uso.....	50
Ilustración 11. Matriz de trazabilidad (Detalle 1)	76
Ilustración 12. Matriz de trazabilidad (Detalle 2)	76
Ilustración 13. Matriz de trazabilidad (Detalle 3)	77
Ilustración 14. Framework Web2py.....	78
Ilustración 15. Estructura Web2py	79
Ilustración 16. Entorno operacional	80
Ilustración 17. Modelo de datos	81
Ilustración 18. Modelo de datos Twitter	82
Ilustración 19. Entidades de twitter.....	84
Ilustración 20. Modelo de datos propio	85
Ilustración 21. Modelo de datos Web2py.....	88
Ilustración 22. Flow de la aplicación	90
Ilustración 23. Prototipo a bajo nivel, Pantalla de registro	91
Ilustración 24. Prototipo a bajo nivel, Pantalla de Login	91
Ilustración 25.. Prototipo a bajo nivel, recuperación de contraseña.....	92
Ilustración 26. Prototipo a bajo nivel, menú principal.....	92
Ilustración 27. Prototipo a bajo nivel, Pantalla búsqueda de tweets	93
Ilustración 28.. Prototipo a bajo nivel, Panel de análisis	94
Ilustración 29. Prototipo a bajo nivel, Pantalla de cuentas relevantes	95

Ilustración 30.. Prototipo a bajo nivel, Base de datos - cuentas relevantes.....	96
Ilustración 31.. Prototipo a bajo nivel, Base de datos – búsqueda de tweets.....	96
Ilustración 32.. Prototipo a bajo nivel, Base de datos - búsqueda de agrupaciones de tweets	97
Ilustración 33. Diseño del sistema, Moodboard	99
Ilustración 34. Tecnologías utilizadas, justificación de elección Vue.js [33].....	101
Ilustración 35. Resultados, registro de usuarios	104
Ilustración 36. Resultados, registro de usuarios (Mobile)	104
Ilustración 37. Resultados, pantalla de login	105
Ilustración 38. Resultados, pantalla de login (Mobile)	105
Ilustración 39. Resultados, pantalla de búsqueda de tweets	106
Ilustración 40. Resultados, resultados de la búsqueda por usuario	107
Ilustración 41. Resultados, base de datos – tweets.....	108
Ilustración 42. Resultados, base de datos - detalle de búsqueda.....	108
Ilustración 43. Panel de análisis	109
Ilustración 44. Panel de análisis - Resultados del análisis.....	111
Ilustración 45. Resultados de análisis	111
Ilustración 46. Encontrar tweets similares	112
Ilustración 47. Agrupación de tweets	113
Ilustración 48. Visualización de tweets agregados	114
Ilustración 49. Base de datos, consulta de tweets agregados	115
Ilustración 50. Pantalla de control de bots	116

ÍNDICE DE TABLAS

Tabla 1. Gestión de proyecto, asignación de recursos (%)	31
Tabla 2. Gestión de proyecto, tabla de costes salariales	33
Tabla 3. Gestión de proyecto, tabla de recursos tecnológicos	33
Tabla 4. Gestión de proyecto, tabla de costes totales.....	34
Tabla 5 Tabla de modelo de datos - Status.....	83
Tabla 6. Tabla de modelo de datos - Usuario	84
Tabla 7. Modelo de datos propio - entidad tweet.....	86
Tabla 8. Modelo de datos propio - agrupación de tweets.....	88

1. RESUMEN Y ESTRUCTURA

1.1. Resumen ejecutivo

Los bulos y las noticias falsas son un problema han existido desde tiempos antiguos, pero la era digital y la normalización del uso de las redes sociales han magnificado el alcance y los efectos del problema. Conceptos como la viralización o el anonimato en las redes sociales han potenciado el impacto de las noticias falsas, haciendo que agencias de noticias, empresas y gobiernos comiencen a tomar medidas para tratar de analizar o solucionar este problema.

Este proyecto trata de proporcionar una herramienta que ayude a todas aquellas personas o entidades que decidan afrontar este problema mediante el desarrollo de una aplicación web accesible para cualquiera que permita analizar la forma en la que se difunden las noticias en las redes sociales, centrando su funcionalidad en la red social Twitter. Para ello se ha construido una aplicación web que permite analizar la difusión de noticias falsas mediante herramientas de visualización de datos, además de la detección de cuentas bot, la agrupación automática de tweets relacionados con una misma noticia o el envío de respuestas a un número elevado de tweets de forma simultánea. La aplicación construida sirve como un mínimo producto viable que permite validar la viabilidad de la idea, de cara a hacer pruebas en un entorno real, recibir feedback y orientar las futuras versiones de la herramienta.

El presente documento trata de exponer el problema y su contexto, así como el proceso y la documentación técnica necesaria que pueda ayudar a reproducir dicho sistema, exponiendo al final los resultados de la aplicación construida, las posibles vías de mejora y oportunidades de explotación de la misma.

1.2. Keywords

Fake News, Aplicación Web, Herramienta de análisis, Noticias falsas, Visualización de datos.

1.3. Objetivos y motivación

El principal objetivo de este proyecto consiste en crear un mínimo producto viable en forma de aplicación web, que permita ayudar a aquellas personas y entidades que se dedican a la lucha contra las noticias falsas. Además, con el proyecto se pretende demostrar los conocimientos adquiridos en diferentes asignaturas relacionadas con el desarrollo de software y la gestión de proyectos.

La motivación de construir esta herramienta web se debe principalmente a tres aspectos: en primer lugar, a la importancia que ha cobrado hoy en día el problema de las noticias falsas, alcanzando las diferentes esferas de la sociedad y de la política. En segundo lugar, por el deseo de crear un sistema que pueda tener un impacto positivo para la sociedad. Por último, por el aprendizaje que supone el uso de tecnologías aprendidas en la carrera, así como explorar en otras tecnologías y frameworks que no se han usado durante el grado, como son Vue.js o Web2py.

A lo largo del desarrollo del proyecto, el tema ha ido cobrando cada vez más relevancia, debido a la situación excepcional que ha supuesto el Covid-19, y al aumento considerable de uso de las redes sociales y de la difusión de noticias falsas durante el periodo de confinamiento.

1.4. Estructura

El proyecto consta de nueve secciones que recogen las diferentes partes del proceso que ha seguido el proyecto con el objetivo de que el sistema que se desea construir pueda ser replicado teniendo esta documentación como base. Las secciones mencionadas son las siguientes:

En primer lugar, se presenta la introducción del proyecto. En esta sección se desea reflejar el problema que suponen las fake news, y la importancia de contar con las herramientas adecuadas para combatirlas. Es aquí donde se introduce la propuesta del sistema que se quiere construir, así como permitir al lector hacerse una idea de las capacidades y alcance del sistema y de los stakeholders interesados.

Una vez introducido el problema y la solución que se desea construir, se sitúa al lector en el contexto del problema, con un análisis de la situación actual, del contexto socio económico, del marco regulador y del estado del arte.

A continuación, se presenta la sección que define la planificación y gestión del proyecto, donde se estiman los costes, recursos y tiempo necesario para completar el proyecto, junto con los pasos necesarios para replicar el sistema usando el conocido diagrama de Gantt.

En la quinta sección, se expone el análisis técnico del sistema, donde se recogen la especificación de requisitos y de casos de uso que definen el sistema.

Posteriormente, se expone el diseño del sistema, que recoge la arquitectura del mismo, así como la definición del modelo de datos.

Una vez se cuenta con el diseño del sistema, se expone el diseño de la interfaz del sistema, junto con uno de los prototipos a bajo nivel usados durante el proceso, que permiten al lector hacerse una idea mucho más visual de cómo será el estado final del sistema, así como recibir feedback de calidad en la fase previa a la construcción.

Para terminar, se exponen las dos últimas secciones que recogen los resultados del proyecto, así como las posibles futuras vías de exploración y las conclusiones a las que el autor ha llegado tras la finalización del proyecto.

1.5. Glosario

Ataque por inyección SQL: tipo de ataque informático que consiste en una ejecución de comandos en una base de datos de forma que aprovecha las vulnerabilidades de la misma para extraer, modificar o eliminar información.

Base de datos: Conjunto estructurado de datos que se almacenan en un sistema.

Back-end: Parte de la aplicación que se encarga de establecer y controlar la lógica del servidor, la gestión de los datos y de atender a las llamadas recibidas desde la parte del cliente.

Caso de uso: Herramienta que se usa para describir las posibilidades de un sistema, así como la forma en la que los actores de un sistema interactúan con el mismo.

Cross Site Scripting (XSS o Ataque por secuencia de comandos), tipo de ataque informático que aprovecha vulnerabilidades de un sitio web para ejecutar código como parte del sistema.

Diagrama de Gantt: Diagrama utilizado en la gestión de proyectos que permite visualizar cada una de las tareas o módulos del proyecto, su duración, secuencia y otros datos opcionales como recursos asignados o tipología de la tarea.

Fake New: Noticia falsa

Framework: Conjunto de herramientas, conceptos y técnicas que crean un entorno estandarizado de trabajo para el desarrollo de software.

Front-end: Parte de la aplicación web que se encarga de gestionar la interacción con el usuario

MVC: (Modelo-Vista-Controlador) Modelo de arquitectura de software en el que el sistema está desagregado en el que la lógica, la interfaz y los datos están separados en los tres elementos mencionados del sistema.

Requisito: Característica que un sistema debe poseer.

Stakeholders: Cada uno de los agentes interesados de manera directa o indirecta en un proyecto o empresa.

Sprint: Intervalo de tiempo utilizado en metodologías ágiles de desarrollo de software para el desarrollo del producto

url shortener: Técnica que permite acortar una url a una url con un número reducido de caracteres.

Web scrapping: Técnica que permite extraer datos de una página web de forma automatizada.

1.6. Lista de Acrónimos

API: Application programming interface.

BBDD: Base de Datos

GDPR: General Data Protection Regulation

IEEE: Institute of Electrical and Electronics Engineers

MVC: Modelo – Vista – Controlador

MVP: Mínimo Producto Viable

SQL: Structured Query Language

RF: Requisito Funcional

RGPD: Reglamento General de Protección de Datos

RNF: Requisito No Funcional

UI: User interface

UX: User experience

2. INTRODUCCIÓN

2.1. Introducción

El conjunto de factores como la digitalización, el acceso a la información, las redes sociales y las nuevas tecnologías, han creado un aumento considerable del consumo de noticias online para cualquier usuario con acceso a internet. Si bien esto es sin duda un aspecto positivo para el usuario consumidor de información online, la digitalización ha traído o adaptado algunos nuevos problemas a la sociedad digital, como son el exceso de la información o las noticias falsas.

Se conoce como fake new o noticia falsa a aquellas noticias o elementos de información que no son veraces o que tienen tal grado de alteración que pueden considerarse como manipulaciones. Si bien el origen del término era aplicado expresamente a las noticias, hoy en día el término se usa de forma popular para referirse a elementos de información, cadenas de mensajes o frases que han sido alteradas, sacadas de contexto o son directamente falsas, lo que comúnmente se conoce como bulo.

Las redes sociales hoy en día tienen un protagonismo esencial en lo que respecta a la difusión de la información, un estudio realizado por el centro de investigación Pew Research Center [1] indica que en 2017 cerca del 70% de los adultos estadounidenses han hecho uso ocasional de las redes sociales como medio para obtener información, y un 20% lo hace a menudo. El mismo estudio indica que desde 2010, año que supuso un punto de inflexión a este respecto, el número de personas que usan las redes sociales como medio de acceso a la información supera al número de personas que acceden a la prensa escrita, haciendo que un 20% de los usuarios acceden a la información a través de canales como Facebook o Twitter, por encima del 16% que lo hace a través de la prensa escrita [2].

2.2. Relevancia de la tecnología en el problema de las fake news.

Cuando un usuario o un medio difunde una noticia en una red social, ésta tiene la capacidad de expandirse de forma masiva y alcanzar millones de visualizaciones en apenas unas horas. Cuando la noticia se difunde de forma especialmente acelerada en comparación con otras noticias similares, se denomina una noticia viral.

Al producirse la combinación de los efectos mencionados anteriormente (noticias falsas y virales) se produce un enorme daño para la sociedad, puesto que, si bien la difusión de la noticia inicial es

tremendamente rápida y el alcance de la misma es amplio, una posible corrección posterior no es necesariamente igual de viral ni tendrá el mismo alcance.

Es importante destacar que no todas las noticias falsas que se comparten en las redes sociales tienen intención dañina o existe un conocimiento de su falsedad por parte del usuario que las comparte. Muchas veces las noticias son tan llamativas o tienen titulares capciosos que enseguida llaman la atención de los usuarios e inducen a los mismos a compartir la noticia, sin prestar especialmente atención a las fuentes, al medio o la persona que los difunde, o a posibles manipulaciones que pueda contener. Este hecho abre una ventana de oportunidad a la solución parcial de este problema mediante una mayor educación de los usuarios y mediante el uso de nuevas tecnologías para el control de la difusión de fake news.

Por otro lado, aquellos usuarios que tengan como objetivo la difusión de noticias falsas de forma consciente, pueden protegerse en el anonimato que las redes sociales ofrecen, o aprovecharse de las posibilidades de automatización que algunas permiten. Una de las principales redes sociales que reúne estas características es Twitter, que proporciona a sus usuarios la capacidad del anonimato y permite a través del uso de sus herramientas para desarrolladores, la conexión de sistemas automáticos o bots a la red social.

Para conocer en detalle la magnitud del problema que se presenta, algunos estudios como la fundación Knight Foundation en 2018 [3] que recoge datos de 10 millones de tuits, concluye que en la red social se publican diariamente más de un millón de tuits falsos, buena parte de ellos procedentes de cuentas bots. Otro estudio llevado a cabo por Vosoughi et al. [4] publicado en la prestigiosa revista Science en base a un estudio sobre 126.000 noticias y más de 4,5 millones de retuits establece que las noticias falsas tienen una probabilidad de viralización de entre 10 y 20% más rápido que los hechos. Como aspecto relevante, el término “fake news” ha sido escogido como término del año 2017 por la famosa editorial Collins debido al espectacular aumento en su uso en relación a años anteriores.

En las sociedades avanzadas actuales, se reconoce como un derecho básico la libertad de expresión y de información, sin embargo, en ocasiones los creadores de manipulaciones, noticias tendenciosas o de noticias falsas se amparan en ese derecho para proteger la producción de las mismas. A pesar de lo interesante que resulta ese debate, no es objeto de este proyecto en entrar a analizar si se debiese controlar o limitar ese derecho para evitar el problema que se menciona, por el contrario este proyecto lo que pretende es crear una herramienta tecnológica que facilite a aquellos que lo deseen el estudio de la difusión de las mismas de una forma más sencilla.

2.3. Propuesta de solución

Una vez planteado el problema, se deben plantear soluciones que permitan conocer el impacto del mismo y potencialmente ayudar a contener su alcance.

Para ayudar al análisis y la resolución del problema mencionado, se desea aprovechar las ventajas de la tecnología y las capacidades que ofrece la red social Twitter, para desarrollar una herramienta que permita a un usuario observar la difusión de las noticias en la red social y obtener datos de una manera sencilla. El objetivo es democratizar en cierto sentido la capacidad de análisis, permitiendo el acceso a los datos a cualquier usuario sin necesidad de tener conocimientos técnicos.

En concreto, se desea desarrollar una aplicación web que permita el análisis de la difusión de tweets y noticias en la red social Twitter y el seguimiento del alcance de las mismas de una forma sencilla o automatizada. Además, se desea incluir en la herramienta la opción de análisis de bots debido a la repercusión que suelen tener estos en la difusión de noticias falsas.

2.4. Descripción y alcance del sistema

El alcance de este proyecto está limitado al desarrollo de un mínimo producto viable (MVP) que contenga las principales funcionalidades del sistema, permitiendo de validar las posibilidades de explotación de la idea, para en caso de éxito crear futuras versiones con mejoras de las funcionalidades que se exponen a continuación o la inclusión de nuevas características.

La motivación que lleva a enfocar el proyecto en la red social Twitter se debe a varias razones: en primer lugar, porque actualmente cuenta con más de 152 millones de usuarios [5], en segundo lugar porque es una red habitual en la que se comparten noticias y cuenta con un gran volumen de actividad y de interacciones diarias. Por último, porque cuenta con una API para desarrolladores que permite obtener datos de una forma automatizada y estructurada.

Estas razones convierten a Twitter en una red social ideal para el desarrollo del proyecto. Una vez esté en funcionamiento se puede plantear el uso de otras redes sociales que también cuenten con API para desarrolladores y ampliar el alcance.

2.5. Capacidades generales

A continuación, se presentan de forma general las principales funcionalidades que se desea que el sistema posea.

Para poder analizar y visualizar la difusión de noticias, la primera funcionalidad que se plantea del sistema consiste en la búsqueda de tweets para su posterior análisis. Para obtener dichos tweets, el sistema deberá permitir la búsqueda al menos por dos parámetros: en primer lugar, la opción de buscar los tweets producidos por un usuario en concreto; y por otro lado la opción de la búsqueda de tweets que contengan palabras clave, frases o noticias. Los tweets encontrados a través de estas dos vías deberán mostrarse a través de la interfaz de la aplicación web, mostrando datos relevantes como el contenido del tweet, el usuario del que procede y la fecha.

La segunda función consiste en guardar aquellos tweets que el usuario considere relevantes para poder hacer un estudio posterior con respecto al mismo. Los tweets guardados se almacenarán en una base de datos para que puedan ser recuperados en cualquier momento por el usuario. Del mismo modo existirá la posibilidad de eliminar los mismos de la base de datos si el usuario lo desea.

En tercer lugar, a partir de un tweet guardado, el usuario podrá analizar de qué forma se ha ido difundiendo en la red social. Gracias al uso de herramientas o librerías de visualización de datos es posible reflejar esa difusión exponiendo gráficas y datos relevantes, como el número de retuits que ha alcanzado, usuarios que han ayudado a la difusión de la misma, gráficas de difusión temporal o el alcance potencial de lectores a partir del número de seguidores de los usuarios que han retuiteado.

Con las funcionalidades mencionadas hasta el momento, pueden parecer que se está asumiendo que cualquier tweet contiene necesariamente una noticia dentro, o que un tweet puede ser la única fuente de una noticia en la red social, pero sabemos que esto no es necesariamente así. Es por ello que también se plantean en el sistema las siguientes funcionalidades:

Detección de noticias en un tweet. Esta funcionalidad permitirá detectar si un tweet concreto contiene una noticia oficial embebida en el mismo. Esto se puede detectar gracias si el tweet contiene una url a la fuente inicial de la noticia (ej. publicación de un periódico). Actualmente la red social tiene una limitación al número de caracteres de un tweet, por lo que cuando se postea una noticia en la red, ésta automáticamente utiliza un acortador de urls que permite reducir notablemente la longitud de caracteres del propio link. Esto plantea un nuevo problema, ya que dos “short urls” pueden referirse a la

misma noticia pese a no ser iguales. Para ello se plantea la posibilidad de que el sistema utilice técnicas de web scraping, que son un conjunto de técnicas que permite extraer información de páginas web de forma automatizada (como por ejemplo el título de la noticia contenida en esa url).

Búsqueda de tweets similares: esta funcionalidad permitirá buscar en Twitter otros tweets similares a los del tweet estudiado. Esta similitud podrá hacerse de acuerdo a al menos uno de los siguientes parámetros (pudiendo ampliarse a varios de ellos en futuras versiones): tweets que contengan la misma noticia, funcionalidad que se puede conseguir gracias a las técnicas de web scraping mencionadas en el párrafo anterior; tweets que contengan texto similar, utilizando técnicas de procesamiento del lenguaje natural para hacer una limpieza del propio tweet, extraer la información relevante y obtener tweets similares; y finalmente la obtención de tweets similares que contengan los mismos hashtags o similares.

Agrupación de tweets: para poder estudiar tweets que sean similares, deberá ser posible la agrupación de los mismos en una entidad que permita su estudio de forma conjunta y agregada. Dicha agrupación debe ser almacenada en una base de datos, y debe ser posible en cualquier momento la inclusión de nuevos tweets similares que surjan, o la actualización o eliminación de los mismos o de la propia entidad agrupación.

Detección de bots: si bien un bot no es un elemento necesariamente relacionado con fake news, en la práctica el uso de los mismos está extendido en la red social siendo en muchas ocasiones uno de sus principales usos el de aumentar el alcance de la difusión de contenido falso o con intenciones dañinas. Dado que actualmente existen proyectos que a partir de un perfil de Twitter estudian la probabilidad de que éste sea un bot en función de multitud de parámetros y variables, se plantea la inclusión de esta funcionalidad en la aplicación web mediante el uso de una API a estos sistemas que ya están contruidos.

Alertar a los usuarios: Para ayudar a contener la difusión de las noticias falsas en la red social, el sistema deberá facilitar una interfaz que permita al usuario de la herramienta responder a los tweets obtenidos. Esto permitirá alertar al usuario de que se está difundiendo una noticia falsa, con la posibilidad de incluir las fuentes que se consideren relevantes para desmentir la noticia.

Para poder hacer un seguimiento más cercano a aquellas cuentas que el usuario considere relevantes o que son objeto de estudio continuo, el sistema contendrá una vista que permita ver los últimos tweets que han posteoado dichas cuentas, además de las opciones necesarias para agregar o eliminar cuentas de la lista de seguimiento.

Si bien algunas de las características presentadas como la obtención de tweets similares o las alertas a los usuarios pueden ayudar en cierta forma al análisis de la difusión de un tweet, estas características pueden ser especialmente potentes si se consiguen automatizar para que se realicen de forma autónoma. Como característica interesante del sistema se desea explorar la forma de automatizar el proceso de detección de tweets similares a la noticia que se está analizando, y el sistema de alerta a los usuarios

Además de las funcionalidades específicas del sistema que se desea construir, se desean incluir también funcionalidades básicas de cualquier aplicación web, como es el hecho de un sistema de registro y login, necesario para el acceso y uso del sistema, con su correspondiente base de datos de los usuarios.

NOTA: Hasta el momento, es posible que el lector se plantee una pregunta esencial: si se trata de un sistema para visualizar la difusión de fake news, ¿de qué forma se va a detectar si una noticia es falsa o no? Ante esta pregunta, es importante mencionar que el objetivo último del trabajo es crear una aplicación web funcional y que el tema de las fake news se ha escogido por ser de interés popular, pero no es el objetivo último del trabajo el de detectar si una noticia es falsa o no sino crear una herramienta que facilite en cierto sentido su detección. Si bien se ha planteado en varias ocasiones a lo largo del proyecto la posibilidad de incluir esta funcionalidad, cabe destacar que hoy en día existen multitud de investigadores experimentados que están estudiando la posibilidad de detectar noticias falsas de forma automatizada (por análisis de patrones y técnicas de Machine Learning entre otros) y hoy en día no existen aún soluciones que se consideren lo suficientemente completas o automatizadas como para poder incluirlas en esta aplicación web. Además, el estudio de noticias falsas con altos niveles de certeza es un campo de análisis con la suficiente extensión en sí misma como para un doctorado, por lo que considero que a pesar de ser un campo muy interesante y que no descarto en algún momento investigar e incluir en la aplicación, por el momento se aleja del alcance del mismo.

2.6. Identificación de los interesados del sistema (stakeholders)

A continuación, se presentan aquellos grupos de personas o entidades que tienen relación con el sistema, incluyendo entre ellos aquellos que son potenciales clientes o usuarios del mismo, si bien aún no hay establecida ningún tipo de relación contractual y podrían terminar excluyéndose de la lista de stakeholders:

El equipo de desarrollo. Para la creación de la prueba de concepto que se expone en este documento, yo he sido la única persona encargada del desarrollo del mismo. Pero el presente documento tiene como una de sus finalidades la de exponer los recursos necesarios para que otro equipo pueda desarrollar y entender el sistema, por lo que considero que si se desea reproducir este proyecto se debe contar con un equipo formado por:

- Jefe de proyecto: es la persona considerada como principalmente del proyecto. Su principal responsabilidad es la planificación del mismo, la asignación de recursos y hacer el seguimiento controlando que los plazos de entrega se cumplan y que las funcionalidades descritas se desarrollen con la calidad establecida en los términos del proyecto.
- Desarrollador front-end: es el integrante del equipo que diseña y construye la forma en la que los usuarios interactúan con un sistema, así como la experiencia de los mismos puede ser determinante para el éxito de un proyecto. Entre sus principales funciones están la de crear las interfaces con las que el cliente interactúa, y crear una experiencia de usuario que sea atractiva y funcional.
- Desarrollador back-end: persona encargada del desarrollo de la parte del sistema relacionada con el servidor, con la base de datos y con conexiones con subsistemas externos entre otras cosas

Usuarios del sistema: el usuario es la persona última que interacciona con el sistema, será la que se dedique a explotar el mismo para el análisis de la difusión de noticias falsas y de sus otras funcionalidades.

Agencias de noticias: como se mencionaba en la introducción, existen agencias como Newtral o Maldita que se dedican expresamente al análisis de noticias y a desmentir bulos. Estas agencias pueden ser uno de los principales stakeholders del sistema, puesto que los objetivos del sistema están alineados con los objetivos de estas agencias de noticias, por lo que pueden beneficiarse de una aplicación de estas características.

Gobiernos y entes públicos: el interés de gobiernos y países con respecto a la lucha contra las noticias falsas y los bulos, pueden hacer de estos un importante stakeholder, especialmente a la hora de conceder ayudas con fondos públicos para sistemas que proporcionen soluciones a este problema.

Comunidad investigadora: cada vez existe un mayor número de investigaciones y estudios que giran en torno a la problemática que se presenta, por lo que la comunidad investigadora podría ser un especial interesado en un sistema de estas características.

Empresas y entes privados: a menudo, las empresas ven perjudicada su imagen de marca por ciertos bulos que se difunden rápidamente por las redes, por lo que es frecuente inversión por parte de empresas en plataformas de escuchas sociales para detectar y gestionar crisis de este tipo en las redes sociales.

La sociedad: la sociedad en sí misma es la que sufre en su mayor medida los efectos de las noticias falsas, por lo que cualquier solución que avance en este sentido tendrá un impacto positivo en la sociedad.

Compañías de redes sociales: las redes sociales son hoy en día uno de los principales medios de distribución de noticias, las compañías tecnológicas son conscientes de la importancia de la lucha contra las noticias falsas, como se expondrá más adelante en las siguientes secciones.

3. CONTEXTO Y ESTADO DEL ARTE

3.1. Valoración de la situación actual

El presente proyecto ha sido realizado en una de las situaciones más excepcionales de los últimos años a nivel Mundial: el momento en el que la pandemia del Covid-19 se ha esparcido por todo el mundo, haciendo que los países se enfrenten a situaciones excepcionales llegando a decretar el Estado de Alarma en España y sus análogos en otros países.

Esta situación es completamente relevante en este proyecto debido principalmente a los siguientes motivos:

En primer lugar, el estado de Alarma ha supuesto una situación completamente anómala por la cual el gobierno ha decretado un estado de confinamiento en el que la mayoría de la población debe permanecer en sus casas evitando al máximo el contacto con el resto de las personas. Con ello se pretende minimizar el riesgo de contagio entre la población. El hecho de que la gente tenga limitaciones para salir de sus viviendas ha supuesto un aumento considerable en el uso de la tecnología y las redes sociales como forma de relacionarse e interactuar con los demás [6]. Este aumento también ha sido notable en la red social objeto de este trabajo: Twitter, tal y como indica Pérez en su artículo [7].

En segundo lugar, esta situación anómala ha creado nuevos temas de interés, de información y de noticias que se exponen en Twitter y en otras redes sociales que los usuarios consumen y sobre las que se interacciona de forma masiva. Como se mencionaba en la introducción, el acceso masivo a la información y la sobreinformación crea una situación perfecta para la difusión de noticias falsas y bulos, que a menudo son difíciles de distinguir de las noticias reales y que se transmiten con facilidad en las redes sociales.

En tercer lugar, en España el problema de las noticias falsas ha alcanzado la esfera política, dado que partidos de todo el espectro ideológico se han acusado mutuamente de transmitir bulos, o del uso de agencias de medios privadas para la censura y el control. Como ejemplo, se exponen dos noticias. La primera noticia [8] es referente al gobierno, en la que varios medios publican una rueda de prensa donde el director de la guardia civil indica que se usan recursos del estado para la lucha contra las noticias falsas y "minimizar el clima contrario a la gestión de crisis por parte del Gobierno".

En la segunda noticia [9] se refleja la reacción a la noticia anterior por parte de partidos de la oposición, que acusan al gobierno del uso de recursos del estado para censurar opiniones contrarias, y del uso de agencias de medios privados como Newtral o Maldita.es con medios afines al gobierno.

Si bien ambas noticias han sido matizadas por cada una de las partes con posterioridad (como las declaraciones del ministro de interior Fernando Grande-Marlaska que refleja la Vanguardia [10]), esto no ha evitado que el tema de los bulos y las redes sociales se haya elevado al punto del debate político. Además, en la mencionada rueda de prensa, el Jefe del Estado Mayor de la Guardia Civil menciona la lucha por parte del cuerpo contra los bulos que se emiten de forma constante en las redes sociales.

Sin entrar en valoraciones políticas, estos hechos no dejan duda de la magnitud que tiene el estudio de las noticias falsas en las redes sociales, y de la importancia de encontrar con los medios necesarios para poder realizar dichos estudios y análisis (como pretende este proyecto).

3.2. Contexto socio económico

La pandemia del covid-19 también ha supuesto un cambio del contexto económico a nivel mundial. En la situación previa a la pandemia, el fondo monetario internacional (FMI) exponía previsiones de crecimiento económico de 2.9% a nivel mundial, y de un 2% de crecimiento para España [11]. Tras la llegada de la pandemia el fondo monetario internacional reflejó un cambio radical en sus expectativas de crecimiento a nivel mundial, arrojando una cifra negativa de crecimiento del -3% del PIB mundial.

En cuanto al panorama económico español, a día 15 de abril las expectativas que expone el FMI para España son de una caída del PIB de un 8%, y la subida del paro hasta el 21% [12]. El hecho de que la economía de España se apoye en buena medida en el sector del turismo y de la poca flexibilidad del mercado laboral ha hecho que el FMI tenga previsiones para España más pesimistas que para el resto de la zona Euro [13].

Las cifras mencionadas en los párrafos anteriores están recogidas a mediados de abril de 2020, pero podrían verse agravadas en caso de que la situación de la pandemia se extienda durante varios meses más en el tiempo. En cualquier caso, el panorama económico refleja una crisis económica de repercusión mundial.

¿Significa esto que no hay oportunidad para un proyecto como el que se presenta? Si bien la crisis que se presenta es a nivel mundial, con la llegada de esta crisis pueden presentarse algunas oportunidades especialmente en el sector digital. Algunas revistas como Forbes, anuncian que esta crisis ha facilitado la aceleración de la transformación digital de las empresas [14].

Desde los propios gobiernos y desde la unión europea, se están creando programas y planes de acción para la lucha contra la desinformación, y se presiona a las compañías tecnológicas por hacer más esfuerzos en la lucha contra las noticias falsas [15]. Estos planes de acción están perfectamente definidos por la comisión europea y trabajan en varios aspectos clave, desde la detección de fake news hasta el aumento de la colaboración con plataformas online e industrias para abordar la desinformación. Estos planes de acción suponen una oportunidad para cualquier plataforma, compañía o investigación que aborde el tema de las noticias falsas y la desinformación.

3.3. Marco regulador

En el marco regulador en el que se enmarca este proyecto, cabe distinguir dos elementos importantes: En primer lugar, la ley general de protección de datos [16] (GDPR por sus siglas), y la política de privacidad de Twitter, que regula los aspectos relacionados con la extracción y uso de tweets, y la interacción del sistema con la red social.

3.3.1. Regulación General de Protección de Datos (GDPR)

La GDPR es la normativa creada por la unión europea con el objetivo de regular la protección de datos de sus ciudadanos. Si bien el sistema que se presenta puede ser construido y reproducido con el objetivo de usarse en terceros países, el proyecto en su estatus actual se engloba dentro del entorno universitario de España, por lo que considero que es relevante reflejar la necesidad de cumplir con esta normativa europea. Además, esta propia regulación establece que, aunque el sistema se aloje en un país ajeno a la unión europea, la normativa es de obligado cumplimiento para cualquier empresa que recoge, trata o almacena datos de ciudadanos de origen europeo. Tratar de evitar dicha regulación supone perder oportunidades en un mercado tan importante como el europeo, con el añadido de que esta normativa es una referencia para otros países, que están tomando medidas para aproximarse a esta regulación. Incluso en Estados Unidos donde cada estado tiene sus propias regulaciones con referencia a este aspecto, se están haciendo esfuerzos importantes por aproximar la regulación existente a la mencionada normativa. Un ejemplo claro de ello es el estado de California (estado referente en Estados Unidos especialmente

por la fuerza de su sector tecnológico) que tiene una regulación muy similar a la GDPR: la ley de Privacidad del Consumidor [17] (CCPA según sus siglas en inglés)

Implicaciones de la GDPR:

La regulación general de protección de datos se compone de 99 artículos que engloban las restricciones que se deben seguir las empresas o entidades que manejan datos, con el objetivo de unificar las regulaciones de los países miembros y proteger a sus ciudadanos. Se puede obtener un análisis avanzado sobre esta regulación en el Informe Entendiendo el Reglamento General de Protección de Datos (GDPR) Un breve recorrido por los aspectos clave Deloitte Consultoría de Deloitte [18], pero los aspectos más relevantes de la misma relacionados con este proyecto son:

- Limitación del propósito de almacenado de datos: Las compañías deben almacenar únicamente aquellos datos que son necesarios para su propósito legítimo. En caso de ampliar el propósito y el uso de los mismos se debe notificar al usuario de forma clara y transparente, y será necesario su consentimiento explícito.
- Derecho a la portabilidad de los datos y al olvido: en cualquier momento, un usuario puede ejercer su derecho de pedir a la empresa toda la información que ésta disponga de él/ella, así como de solicitar el borrado de toda la información relativa al mismo salvo que existan razones importantes para mantener los datos de los usuarios (relaciones contractuales en vigor, impagos, etc)
- Consentimiento: se necesita el consentimiento expreso por parte de los usuarios para que las compañías puedan usar los datos con fines distintos a los fines legítimos por los que fueron recogidos. El usuario puede retirar ese consentimiento en cualquier momento.
- Brechas de seguridad: existe la obligación legal por parte de las empresas de transmitir a las autoridades de supervisión en un rango de 72 horas cualquier posible incidencia que viole la protección de datos.
- Análisis de riesgos y medidas de seguridad: se deben aplicar medidas de análisis de riesgo y de seguridad que permitan conocer y controlar el riesgo que puede suponer para los usuarios un uso indebido o una filtración de los mismos [19].

En el caso de reproducirse este proyecto, recomiendo encarecidamente hacer un estudio actualizado de dicha regulación para permanecer en todo momento en perfecto cumplimiento de la normativa.

3.3.2. Política de privacidad de Twitter

La red social Twitter cuenta con una extensa política de privacidad que regula registro y acceso a la red social, así como el uso de la API por parte de desarrolladores externos a través de la cual se obtienen datos.

Twitter expone de forma clara desde el primer momento que se trata de una red social pública. Cuando un usuario postea un tweet o se crea una cuenta, está exponiendo dicha información al público y se asume que el interés del usuario hacer dicha información pública. Por este motivo Twitter permite su visualización al público (independientemente de que sean usuarios de Twitter o no), y permite la obtención de dicha información a través de su API a desarrolladores externos. Esta política está recogida en el artículo 1.2 de su política de privacidad disponible en su web [20].

Este punto de la política de privacidad es un aspecto clave para este proyecto, dado que el sistema gira en torno a la posibilidad de obtener tweets a través de la API de la red social para poder tratarlos y analizarlos con posterioridad.

La red social cuenta también con una política de privacidad distinta creada para aquellos desarrolladores que van a usar la API. Dado que nuestro sistema se basa en dicha API para obtener información, es importante mencionar aquellos aspectos de dicha política que son relevantes para el proyecto.

Uno de los principales aspectos relevantes para el proyecto que menciona la política de privacidad hace referencia al almacenamiento de contenido de Twitter en sistemas externos. En caso de producirse esta situación (como sucede en este proyecto), se debe mantener en todo momento el contenido con el mismo estado que se encuentra en Twitter. Esto significa que, si en algún momento el autor actualiza su estatus a privado, o elimina un tweet, el sistema deberá actualizarse de acuerdo con el mismo, eliminándose del sistema offline si fuera necesario [21].

La política de privacidad también afecta a otra de las características del sistema: la posibilidad de responder a tweets de forma automatizada a través de la plataforma para alertar a un usuario o a un grupo de usuarios que han posteado una noticia de que pueden estar difundiendo una noticia falsa. En

este aspecto la política de privacidad menciona [22] que se permite automatizar el envío de respuestas si se cumplen las siguientes condiciones:

- Previamente al envío de la respuesta, el usuario ha mencionado o indicado claramente la intención de contactar con la cuenta del sistema
- Se provee al usuario de forma clara una forma de hacer opt-out (retirar consentimiento) ya sea mediante una respuesta al tweet o vía mensaje directo
- Sólo se envíe una respuesta por usuario e interacción
- La respuesta se hace al tweet original del autor.

Nota: Para poder cumplir con esta política se debe replantear alguna de las funcionalidades del sistema, cambiando de una respuesta automatizada a varios usuarios, a una respuesta única y personalizada para cada usuario de forma manual (a través de la interfaz de la aplicación) ofreciendo posibilidad para el usuario de hacer opt-out.

El resto de los artículos que menciona la política de privacidad no son aplicantes por el momento para el alcance de este proyecto, ya que se basa principalmente en la recuperación de la información pública mencionada anteriormente, y la única forma de interacción con otros usuarios se menciona en el segundo artículo expuesto anteriormente [20].

3.4. Estado del arte

Durante esta sección se presentarán algunas soluciones que existen a la hora de combatir los bulos centrándose el análisis en soluciones en el uso de la tecnología y las redes sociales.

Soluciones de las grandes compañías tecnológicas

La preocupación por el aumento considerable de noticias falsas ha hecho reaccionar a las principales grandes tecnológicas como Google, Facebook, Twitter o WhatsApp, que han comenzado a tomar algunas medidas con relación a este problema.

La compañía Google, ofrece una serie de herramientas para desarrolladores dedicadas al fact checking. En concreto, cuenta con un buscador de fact checks y una API que permite usar tecnología de Google para verificar si una noticia o sentencia es falsa o no. Para ello, Google tiene establecidos una serie de acuerdos a nivel mundial con agencias de verificación. Dichas agencias a las que Google considera de

confianza introducen en la base de datos noticias que han sido verificadas, de forma que cualquier persona puede usar el buscador de fact checks con respecto a una noticia, y ver si la noticia está calificada como falsa, confusa, parcialmente verdadera o verdadera, y saber qué agencias han verificado esa noticia.

Por el momento, esta herramienta se centra en frases o sentencias cortas y concretas, por lo que aún no tiene la capacidad de cubrir una noticia entera. Además, las verificaciones que se pueden encontrar en esta herramienta están principalmente en inglés.

Facebook también cuenta con una serie de acuerdos agencias de verificación privadas, a través de las cuales es capaz de marcar noticias publicadas en su red como bulos, o reducir notablemente su exposición al público mediante los propios algoritmos de difusión. Además, cuenta con un sistema de reportes, en el que la propia comunidad puede denunciar una noticia. Si una noticia cuenta con varios reportes puede ser revisada por agentes de Facebook y ver reducida notablemente su exposición al público. La compañía también cuenta con equipos internos [23] que se dedican a la investigación, creación y uso de algoritmos para combatir bulos.

WhatsApp, compañía que pertenece al grupo de Facebook ha introducido recientemente una política mediante la cual se limita el número de personas a la que se puede reenviar un mensaje, con el objetivo de ralentizar la viralización del mismo y ayudar a prevenir la difusión de noticias falsas [24].

Twitter, por su parte, también está avanzando en el uso de la tecnología para combatir las noticias falsas. Hace unos meses la compañía anunciaba la adquisición de otras empresas como Fabula AI [25], con el objetivo de usar la inteligencia artificial y el aprendizaje automático para (entre otras cosas) luchar contra la difusión de bulos. Además, también experimentando con una nueva característica [26] que marca visualmente con color naranja aquellos tweets de los que se tiene sospecha de contener una noticia falsa.

Agencias de verificación

En los últimos años, algunos periodistas decidieron agruparse para montar agencias que se dedican expresamente a la verificación de noticias falsas y fact checking. A modo de ejemplo en España, podemos encontrar las agencias de Newtral y de Maldita.es que, junto con otras agencias a nivel internacional, forman parte del IFCN (International Fact-Checking Network), una unidad de entidades que se dedican a la verificación de noticias.

Estas agencias, se dedican a buscar las fuentes de información de noticias que los propios usuarios les hacen llegar a través de las redes o de sus páginas web, y gracias a un equipo de investigadores, periodistas y el uso de la tecnología, investigan y verifican las mismas atendiendo a sus fuentes primarias y a los datos. Además, en muchas ocasiones tienen una labor educativa importante, con la creación de guías y consejos para evitar caer en bulos.

Estas agencias, suelen centrar su actividad en las redes sociales debido a la gran actividad que existe en las mismas y a que son uno de los principales medios de difusión de noticias. Durante el último año 2019, dichas agencias han comenzado a formar parte como verificadores oficiales en los programas especiales en colaboración con compañías tecnológicas mencionados anteriormente.

Herramientas de detección de bots.

El uso de bots en las redes sociales es algo común cuando lo que se desea es automatizar o aumentar el alcance de difusión a la vez que se mantiene el anonimato. Es por ello por lo que los bots en las redes sociales a menudo se relacionan con campañas agresivas de difusión, en muchas ocasiones de noticias falsas.

Existen actualmente algunas herramientas creadas que ayudan a detectar si una cuenta en las redes sociales es un bot o no, mediante el uso de la tecnología para detectar patrones o el uso de inteligencia Artificial y procesamiento del lenguaje natural. Algunos ejemplos de ello son:

Botometer: es una web que cuenta con una API que, mediante el uso de Machine Learning clasifica cuentas de Twitter con una probabilidad de que la cuenta sea bot o no. Además, dispone de una Api para su uso por desarrolladores.

Bot Sentinel: es una herramienta similar a la anterior, que cuenta con una base de datos de cuentas analizadas que los usuarios pueden consultar.

Herramientas de visualización de datos

Las herramientas de visualización de datos son especialmente relevantes a la hora de combatir fake news dado hacen mucho más accesible el análisis y el acceso a la información a personas o entidades menos familiarizadas con la tecnología, centrando sus esfuerzos únicamente en la detección y análisis de fake news, evitando invertir gran cantidad de tiempo y recursos en aspectos más tediosos como la

extracción y limpieza de información. Además, contar con las opciones de visualización adecuadas permite la extracción de conclusiones mucho más exhaustivas en los análisis. Algunos ejemplos de este tipo de aplicaciones son *Ide tweetreach* o *mention map*. Además, existen multitud de compañías que cuentan con herramientas avanzadas propias de análisis de datos, que permiten una visualización avanzada de los datos además de ofrecer al usuario posibilidades de interacción y filtrado de los mismos para un análisis más profundo. Dos ejemplos de ellas son la compañía *Graphext* o la compañía *Graphika*. Ambas cuentan con interesantes análisis de datos en redes sociales.

Existen además una serie de potentes herramientas y plataformas que permiten detectar, automatizar y visualizar el impacto y la difusión en redes sociales, pero suelen estar principalmente orientadas a la imagen de marca. Estas herramientas permiten visualizar y detectar interacciones en la red relacionadas con la marca, visualizar interacciones y tendencias en la red, y tomar acciones automatizadas como responder a mensajes o crear alertas o sistemas de gestión de crisis. Si bien estas herramientas se alejan un poco del objetivo de este proyecto, es cierto que podrían usarse en cierta medida para análisis de difusión y para visualización de datos. Entre las más famosas podemos encontrar *Sprinklr* o *Hootsuite*.

Además de las mencionadas, existen otras potentes herramientas de Inteligencia de Negocios que dedicadas a la visualización de datos de forma muy profesional y avanzada, pero necesitan de una serie de conexiones a datos y configuración previa, o de ingesta manual de datos para poder ofrecer las visualizaciones. Entre las más famosas encontramos *PowerBi*, *Tableau* o *Datorama*.

Herramientas especializadas en fake news

Además de las soluciones mencionadas anteriormente, existen algunas herramientas creadas con el objetivo concreto de combatir las noticias falsas, aunque son herramientas que suelen estar ligadas con la comunidad investigadora y que aún se encuentran en etapas tempranas de desarrollo. Un ejemplo de ello es *Hoaxy*.

Hoaxy es una herramienta de código abierto que combina la parte de visualización de datos con la detección de noticias falsas. La web permite visualizar la difusión de artículos encontrados en Twitter, a partir de cuentas de baja credibilidad o cuentas relacionadas con el fact checking. Esta herramienta en fase beta permite al usuario visualizar de una forma muy básica la difusión de artículos en Twitter, o usar un buscador de tweets a través de su interfaz.

Hoaxy es una herramienta de código abierto de la universidad de Indiana, aunque se encuentra en fase beta y su última actualización data de varios meses atrás, por lo que no parece estar actualizándose.

Otras investigaciones y estudios

Existen numerosos estudios e investigaciones de la comunidad investigadora que se han dedicado al análisis y detección de noticias falsas. Para ello se usan diferentes formas de afrontar el problema que abarcan desde el uso de técnicas y algoritmos de aprendizaje automático, hasta el uso de diferentes enfoques frente a los mismos, como por ejemplo análisis de patrones y tiempos de difusión, análisis de la fiabilidad de la fuente, análisis de contexto o análisis de contenido de la noticia (O una combinación de ellos). En un estudio realizado por Bondielli y Mercelloni [27] se expone de forma extensa los diferentes enfoques actuales, entre los que destacan los siguientes:

Machine Learning

El uso de técnicas de clasificación mediante aprendizaje automático es una de las principales aproximaciones a la hora de enfrentar este problema. Existen algunos experimentos basados en Support Vector Machines que usan análisis semántico, así como análisis de patrones de propagación para detectar rumores, obteniendo resultados de hasta un 91% de fiabilidad como el estudio realizado por Wu et al [28].

Otras investigaciones están centradas en el uso de árboles de decisión que usan un análisis de contexto unido a un análisis de contenido para la evaluación mediante árboles de decisión, llegando a obtener resultados con índices de fiabilidad del 86% [29]. Entre los árboles de decisión, destaca el algoritmo Random Forest, que opera como un conjunto de árboles de decisión a la vez Kwon et al. [30] han implementado un algoritmo Random Forest para clasificación de rumores en grafos de tuits con índices de fiabilidad del 90%

Otros investigadores han afrontado el problema usando regresión logística desde diferentes perspectivas, como el estudio de Momchil et al [31]

Deep Learning

Otra de las formas de afrontar el problema es mediante el uso de Deep Learning. El uso de estos algoritmos ha crecido notablemente en los últimos años, teniendo como principal ventaja frente a algoritmos de aprendizaje automático que estos últimos suelen estar basados en extracción manual de características y elementos, y están sujetos a cierto bias.

Entre los algoritmos de Deep Learning destacan el uso de redes convolucionales y el uso de Redes Neuronales Recurrentes. Algunos estudios realizados en Twitter reportan un porcentaje de precisión del 70%, como el realizado por Chen et All [32].

Principales problemas

En general, la mayoría de los autores e investigadores coinciden en la gran dificultad de obtener altos índices de fiabilidad, debido principalmente a tipo de datasets con el que se trabaja, y a la complejidad de los mismos a la hora de realizar análisis léxicos, sintácticos y de contexto. Por lo que es frecuente que dichos algoritmos reduzcan mucho sus índices de fiabilidad cuando se enfrentan a nuevas fuentes de datos.

Además, la mayoría de los estudios e investigaciones existentes se han realizado en un entorno cerrado, a partir de datasets concretos mediante procesos principalmente manuales o poco automatizados y en inglés. Esto plantea un problema desde el punto de vista tecnológico, aún no hay una verdadera usabilidad de los mismos en tiempo real, por lo que, pese a que los estudios son numerosos y cada vez más prometedores, es necesario integrarlo en herramientas o aplicaciones web como la que se presenta con el objetivo de conseguir una usabilidad real y exponer su utilidad.

El uso de la tecnología para combatir las noticias falsas es un campo de investigación que, debido a su dificultad, se encuentra aún en una etapa inicial o de crecimiento, por lo que existen grandes oportunidades para la creación y explotación de nuevas soluciones y herramientas.

4. GESTIÓN DEL PROYECTO

En esta sección se refleja toda la información relativa a la planificación proyecto, mencionando los recursos necesarios para el proceso de desarrollo del sistema y la forma de gestionarlos.

4.1. Metodología de desarrollo

Como metodología de gestión del proyecto se ha escogido una metodología mixta en la que se combina una primera fase de planificación y documentación más propia de las metodologías en cascada, pero para el desarrollo se van a seguir procesos iterativos basados en Sprint más propios de las metodologías ágiles. Esta elección se justifica en base a los siguientes motivos:

- Las metodologías ágiles siguen un proceso Iterativo e Incremental y tienen como una de sus principales características la priorización del desarrollo y funcionalidades sobre la generación de documentación. Debido a la característica última de este proyecto (un trabajo de fin de grado donde se pretenden demostrar los conocimientos adquiridos en la carrera a través del presente documento), la elección de una metodología que no se apoye en documentación puede impedir demostrar esos conocimientos. Este es el motivo por el que la fase de planificación y documentación del proyecto se hace más extensa, siendo una característica propia de metodologías en cascada
- Para la fase de desarrollo, sin embargo, se considera que el uso de metodologías ágiles y seguir Sprints de desarrollo es la decisión más acertada, puesto que es una forma de marcar objetivos de desarrollo quincenales, se pueden tomar decisiones que permitan incluso cambiar el sistema o añadir nuevas características y no es necesaria la generación de documentación durante el desarrollo, puesto que en este proceso la única persona encargada del desarrollo final ha sido el autor del mismo. Sin embargo, se recomienda la generación de ciertos informes de seguimiento y cierta documentación a aquellas personas o equipos que deseen implementar el proyecto en el futuro.
- Por último, debido al desarrollo de casos de uso, seguir una metodología ágil parece una decisión acertada, ya que en cada sprint se pueden implementar uno o varios casos de uso que tienen un alcance limitado y engloban un conjunto de requisitos definidos y distinguidos por temas.

Debido a las características mencionadas anteriormente, se ha escogido tomar como base la metodología Craig Larman (aunque no en su forma más estricta), que es una metodología iterativa, incremental y basada en casos de uso que cumple con las necesidades expuestas anteriormente, ya que cuenta con una primera fase de planificación y documentación, pero el desarrollo se hace en forma de sprints, donde en cada sprint se implementan nuevos casos de uso.

En esta metodología se distinguen tres etapas fundamentales:

- Fase de planificación: fase en la que se define el sistema, se generan los casos de uso y los requisitos, se establece el alcance del sistema y su arquitectura
- Fase de desarrollo: fase en la que se seguirán una serie de iteraciones llamadas sprints en el que se incluirán características en el sistema de forma incremental
- Fase de deployment o despliegue: fase en la que se instala el sistema y se presenta para su uso

4.2. Planificación del proyecto

A continuación, se muestra la planificación completa del desarrollo del proyecto, a través de un Diagrama de Gantt que refleja las principales tareas a realizar, así como su estimación temporal y su orden dentro del proyecto.

Para poder reflejar y visualizar de manera correcta el diagrama, se ha segmentado en las fases mencionadas anteriormente.

Fase de planificación

Para el proyecto se ha asignado un mes para esta primera fase, que engloba tareas no sólo relacionada con la planificación del proyecto, sino también con la definición del sistema, generando los requisitos, casos de usos y prototipos que permiten concretar el alcance del sistema y de sus características.

En el gráfico quedan reflejados también los diferentes entregables definidos, que marcan pequeños hitos del proyecto. Además, con los diferentes entregables se pretende recoger información de

los diferentes stakeholders interesados, para poder realizar los ajustes necesarios con el feedback obtenido.

NOTA: En las siguientes imágenes sólo se muestra el detalle ampliado. Puede visualizarse el gráfico de forma expandida en el anexo

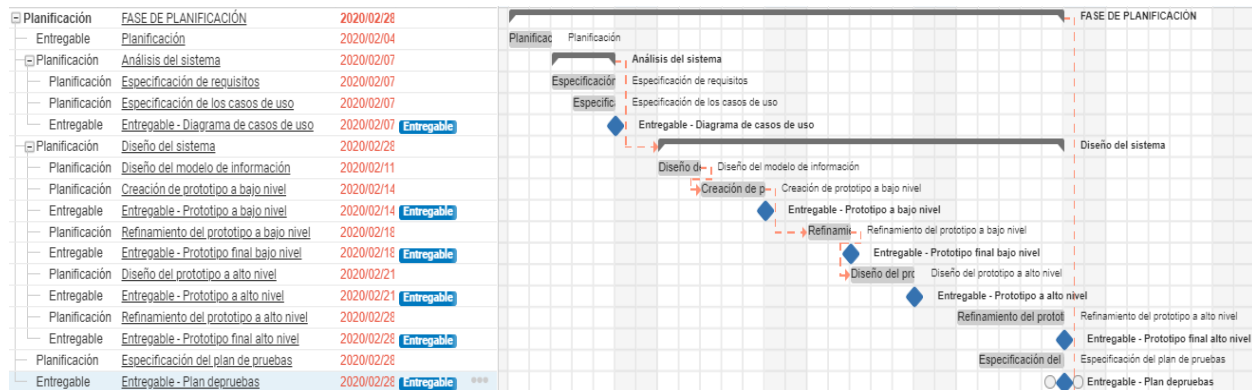


Ilustración 1. Fase de planificación Diagrama de Gantt

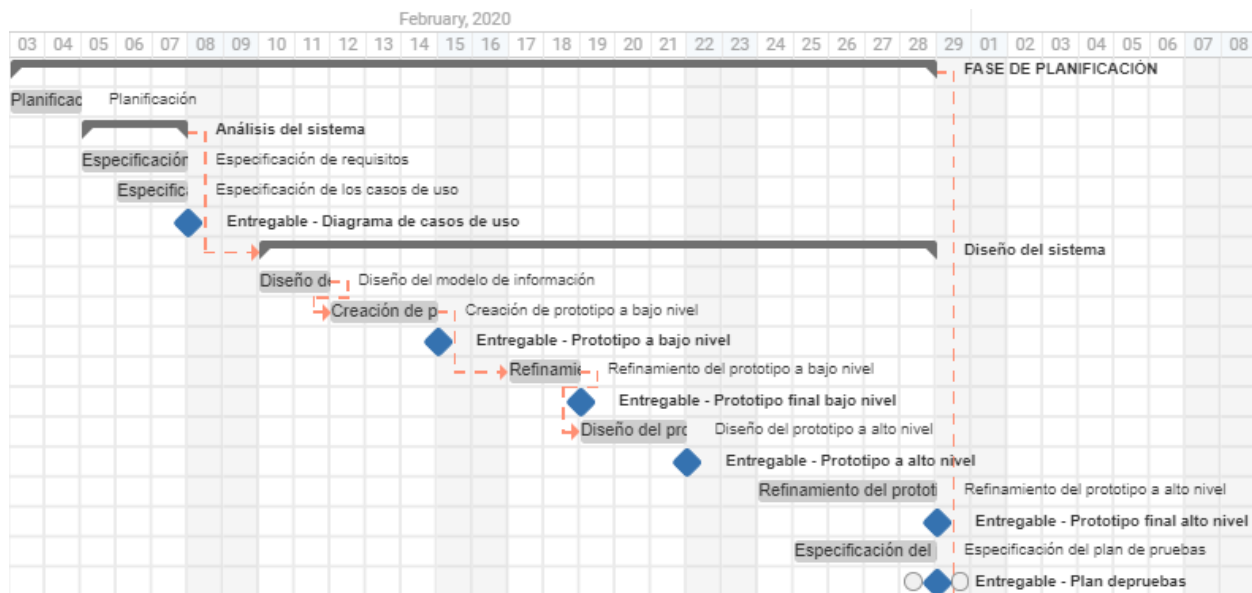


Ilustración 2. Fase de planificación Diagrama de Gantt (Detalle ampliado)

Esta fase tiene una duración de un mes. Como referencia temporal se ha situado el inicio del proyecto en el día 2 de Febrero de 2020, y se reflejan la duración en días de las diferentes tareas

Fase de desarrollo:

A continuación, se muestran los diferentes Sprints que conforman la fase de desarrollo. En cada Sprint, que tiene una duración de dos semanas, se implementarán entre dos y tres casos de uso. El número de casos de uso a incluir en cada Sprint depende de la dificultad de los mismos, y el orden viene marcado por las dependencias reflejadas en el diagrama de casos de uso. Al final de cada Sprint se incluye la fase de pruebas.

Destaca en la primera fase, una parte previa de configuración del proyecto, entornos, librerías necesario para el comienzo del desarrollo.

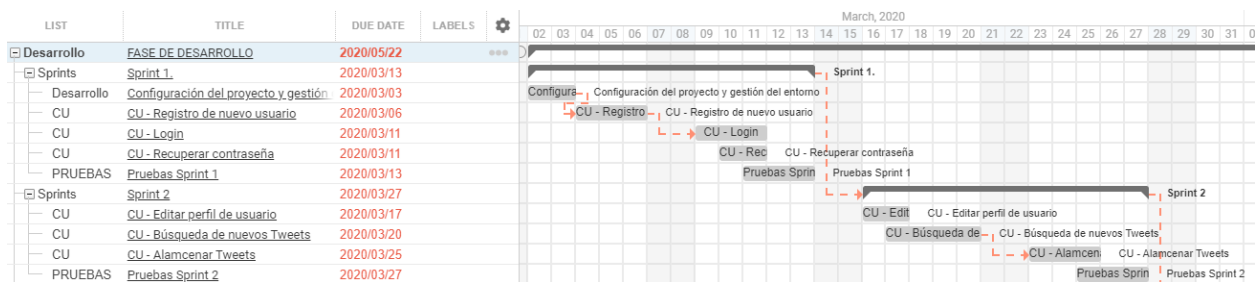


Ilustración 3. Fase de Desarrollo Diagrama de Gantt, Sprint 1 y 2

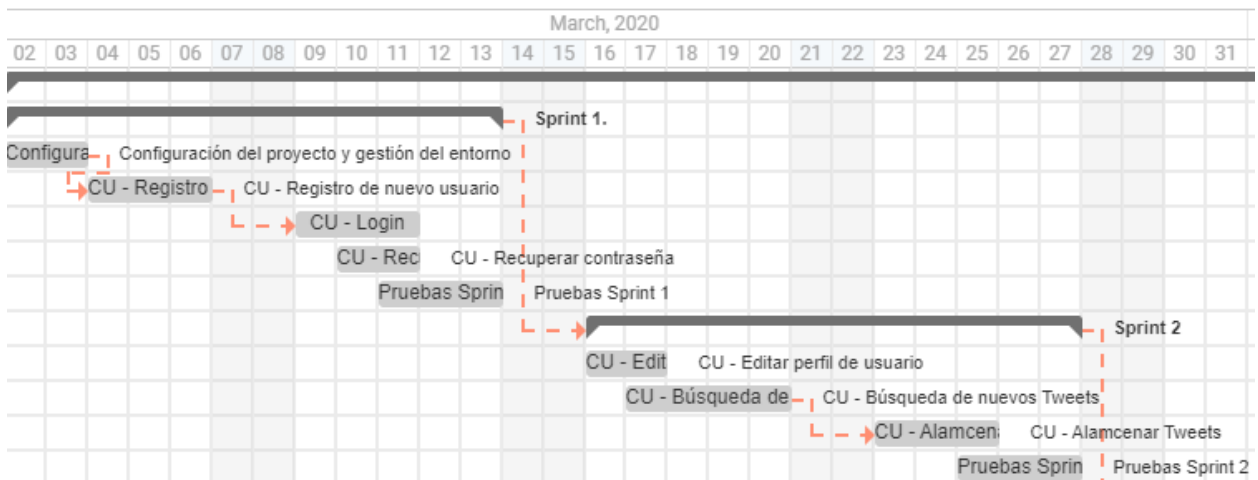


Ilustración 4. Fase de Desarrollo Diagrama de Gantt, Sprint 1 y 2 (Detalle ampliado)

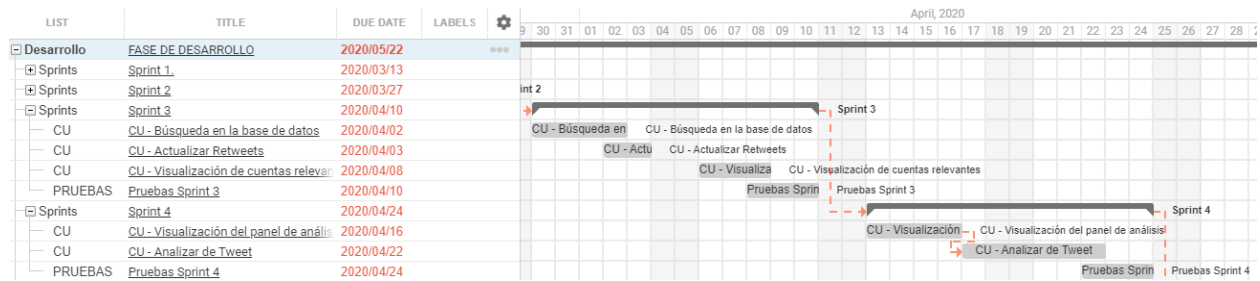


Ilustración 5.Fase de Desarrollo Diagrama de Gantt, Sprint 3 y 4

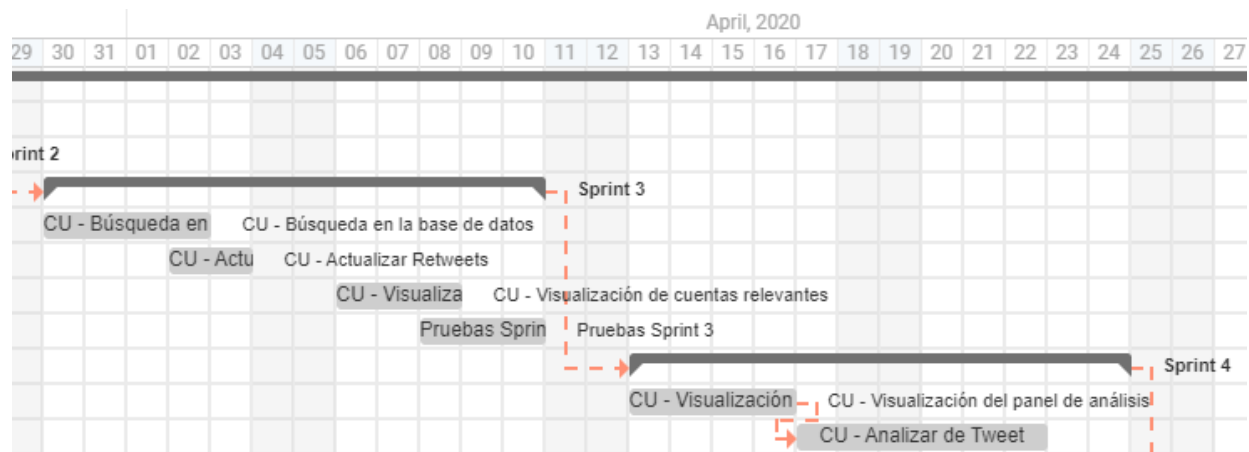


Ilustración 6.Fase de Desarrollo Diagrama de Gantt, Sprint 3 y 4 (Detalle ampliado)

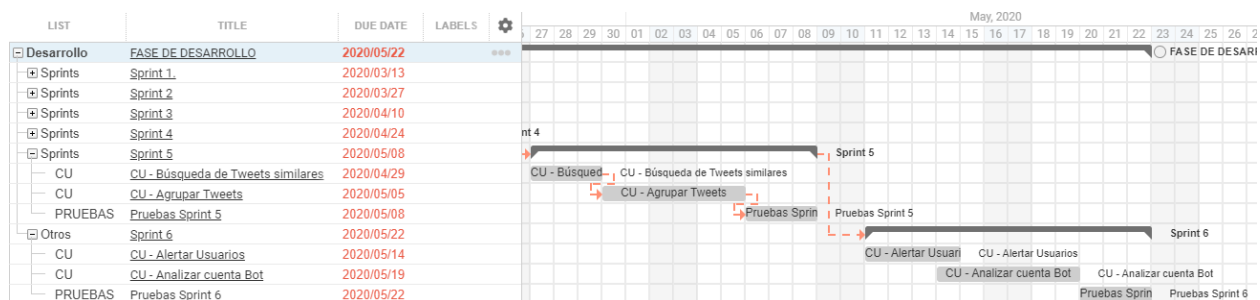


Ilustración 7.Fase de Desarrollo Diagrama de Gantt, Sprint 5 y 6

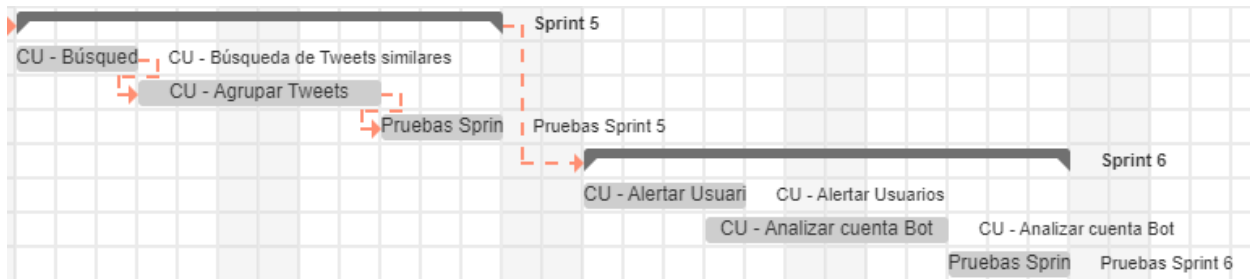


Ilustración 8. Fase de Desarrollo Diagrama de Gantt, Sprint 5 y 6 (Detalle ampliado)

Fase de deployment o despliegue

Tiene una duración de una semana y marca el fin del proyecto. Engloba las tareas de configuración e instalación de la aplicación en los servidores y las pruebas finales.

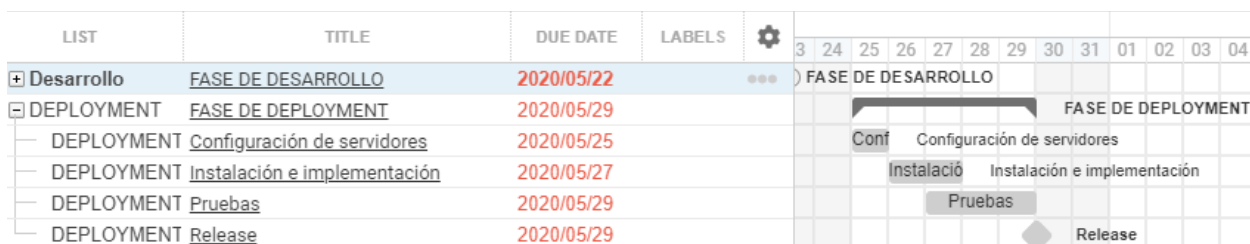


Ilustración 9. Fase de Deployment Diagrama de Gantt

Con esta última fase finaliza el proyecto, teniendo una duración aproximada de 4 meses para la primera versión. Estimación de costes y recursos del proyecto.

4.3. Análisis de recursos del proyecto

Estimación de recursos humanos

Para el desarrollo del proyecto, se recomienda contar con los siguientes perfiles, algunos de los cuales tienen una dedicación parcial al proyecto, por lo que sólo son necesarios en ciertas etapas:

- **Jefe de proyecto:** Esta persona es la encargada de la gestión del proyecto y del equipo, responsable principal de que los entregables se cumplan en tiempo y forma, así como de conseguir y facilitar los recursos necesarios para que los diferentes perfiles puedan cumplir

su trabajo. Además, es la “cabeza visible” del proyecto de cara a los diferentes stakeholders, por lo que cumple la función también de comunicación y negociación con los mismos.

- **Desarrollador front-end:** Este perfil es el responsable de la creación de los diseños de la aplicación y los prototipos en las fases tempranas del proyecto, así como de la codificación de los mismos en las fases de desarrollo*
- **Desarrollador back-end:** Esta persona se encarga del desarrollo de la parte menos visible del sistema, el servidor, así como del desarrollo del código que permite los protocolos de comunicación con el cliente. Además, es el responsable de la codificación y la gestión de la base de datos, siendo así el perfil más técnico del proyecto*.
- **Arquitecto de software:** Este rol, se ve involucrado durante la fase de planificación del proyecto, para elaborar el diseño de los sistemas de información y de las tecnologías a usar.

De cada uno de los perfiles mencionados es exigible el nivel de conocimientos técnicos adecuados para su perfil, así como del conocimiento de los estándares y normas relacionados con el mismo.

Es importante mencionar que los roles expuestos son los que se consideran como óptimos para el alcance de este sistema, pero son adaptables en función del presupuesto y el nivel técnico de los perfiles, además de que pueden ser realizados por la misma persona. A modo de ejemplo, el sistema que se expone en este trabajo ha podido ser implementado por el autor del mismo, Ignacio Laviña, haciendo la función de cada uno de los perfiles mencionados anteriormente.

*Debido al tamaño y a la tipología del proyecto, los **perfiles front-end y back-end** se van a englobar en un único **perfil full-stack**, que engloba los conocimientos de ambos campos. Esto nos permite ahorro en coste de materiales, aunque supone un incremento en costes salariales.

Dedicación al proyecto según perfil:

Perfil	(%) implicación	Etapas de implicación
Jefe de proyecto	30%	Todas las etapas, con especial implicación en la etapa de planificación
Desarrollador full stack	100%	Todas las etapas, dedicación completa
Arquitecto de software	20%	Etapas de planificación

Tabla 1. Gestión de proyecto, asignación de recursos (%)

Recursos tecnológicos

Los diferentes perfiles mencionados, deben contar con el material adecuado para poder trabajar. Para ello se ha asignado un ordenador Mac Book Pro para el desarrollador, y un ordenador Dell para el jefe de proyecto y para el Arquitecto de software.

A pesar de que aquí se reflejan los costes completos, únicamente se asignarán los costes relacionados con la amortización de los mismos en función de la implicación del perfil en el proyecto. Puede verse con mayor detalle en la tabla de costes.



Ordenador Mac Book Pro 16"

Intel Core i7 de seis núcleos a 2,6 GHz
de novena generación, 16 GB de memoria
DDR4 a 2.666 MHz, 512 GB de
almacenamiento SSD, Pantalla Retina de
16 pulgadas

2699€



Ordenador Dell Latitude 7400

i7-10510U de 10.ª generación 8 MB de caché,
hasta 4,9 GHz, 16GB Ram, SSD512, NVIDIA®
GeForce® MX250 pantalla 14 pulgadas

1.089,00 €

Otros recursos y materiales fungibles

Para el desarrollo y despliegue del sistema, es necesario contar con las herramientas de software adecuadas. Si bien la mayoría del software utilizado y expuesto en la lista de tecnologías y software es de licencia gratuita, o tiene planes gratuitos que son suficientes para el alcance del proyecto, para el desarrollo del proyecto se recomienda asignar una parte del presupuesto para hacer frente a estos gastos u otros que pudieran surgir. Para el caso de este proyecto se asigna una cantidad de 200€.

Además, se asigna también 120€ de presupuesto para material de oficina y otros recursos menores.

4.4. Análisis de costes del proyecto

Para la estimación de costes, se van a aplicar solamente aquellos costes calculados con respecto a su porcentaje de amortización, y no por su totalidad. Se entiende que los recursos pueden ser reutilizados en otros proyectos y que los roles involucrados sólo se contabilizan mientras estén involucrados. Los costes reflejados del proyecto pueden variar mucho en función de las características propias de la empresa y de los equipos de desarrollo, por lo que estas cifras deben tomarse como una estimación.

Tabla de costes salariales:

Perfil	Salario bruto	horas anuales	salario por hora	Duración proyecto	% implicación	horas de proyecto	Costes totales
Jefe de proyecto	30,000 €	1760	17.05	608	30%	182	3,109 €
Desarrollador full stack	28,000 €	1760	15.91	608	100%	608	9,673 €
Arquitecto de software	28,000 €	1760	15.91	608	20%	122	1,935 €
TOTALES							14,716 €

Tabla 2. Gestión de proyecto, tabla de costes salariales

Recursos tecnológicos:

Item	Coste	Amortización (total meses)	coste/mes	Amortización (meses proyecto)	Coste total unitario
MacBook Pro	2,077 €	24	87 €	4	346 €
Dell Latitude 7400	1,475 €	24	61 €	2	123 €
Dell Latitude 7400	1,475 €	24	61 €	1	61 €
TOTALES					531 €

Tabla 3. Gestión de proyecto, tabla de recursos tecnológicos

Tabla de Costes totales:

Calculados los costes anteriores, se añaden los costes relativos a los materiales fungibles y a software y licencias. Además, se incluye un margen de riesgo del proyecto.

COSTES TOTALES DEL PROYECTO	
Sueldos empleados	14,716 €
Equipos informáticos	531 €
Materiales fungibles	120
Software y licencias	200
Subtotal	15,567 €
Margen de riesgo (10%)	1,556.69 €
Total	17,124 €

Tabla 4. Gestión de proyecto, tabla de costes totales

5. ANÁLISIS DEL SISTEMA

5.1. Especificación de requisitos

5.1.1. Justificación de la plantilla de requisitos

Para la identificación de requisitos se han usado principalmente tres técnicas: la primera consiste en la investigación de requisitos de otros proyectos de desarrollo de software, tratando de encontrar puntos comunes con el presente proyecto sobre los que extraer y basar algunos de los requisitos de este sistema, entre los proyectos mencionados se encuentran los realizados en asignaturas como Ingeniería del software o Dirección de proyectos de desarrollo de software. Como segunda técnica, se han realizado bocetos a bajo nivel del proyecto, y se han expuesto a otras personas con el objetivo de obtener feedback. Como tercera técnica, se ha seguido un proceso inverso (primero el desarrollo y luego la extracción de algunos requisitos) por lo que estos vinieron después tras el desarrollo de alguna de las funcionalidades del sistema.

Para la redacción de los requisitos, se está basada en el estándar IEEE830, recogiendo de esta plantilla los elementos que se consideran esenciales en este proyecto. Esta elección se basa principalmente en dos motivos: el primer motivo se debe a que es un estándar conocido y usado en otros proyectos con anterioridad. El segundo motivo para la elección de la plantilla IEEE830 se debe a que es un estándar de calidad reconocido en el desarrollo de software.

A continuación, se expone la plantilla utilizada por el estándar:

Requisito	Nombre del requisito
Código	Código identificativo del requisito (RF - XXX RNF - YYY)*
Prioridad	Esencial / Deseado / Opcional
Descripción	Descripción detallada de lo que consiste el requisito

* RF – hace referencia a Requisito Funcional

* RNF – hace referencia a Requisito No Funcional

5.1.2. Justificación de la clasificación de requisitos

Con base en el estándar IEEE830, los requisitos se han clasificado en requisitos funcionales y requisitos no funcionales. A su vez, los requisitos funcionales se han dividido en tipologías propias específicas únicamente para este sistema, que se muestran a continuación:

Requisitos funcionales:

Requisitos de login y registro

- Registro de usuarios
- Recuperación de contraseña
- Login de usuarios
- Fin de sesión
- Visualización de perfil de usuario
- Actualización de perfil de usuario
- Eliminación de perfil de usuario

Requisitos de obtención de Tweets

- Búsqueda de Tweets por tema
- Búsqueda de Tweets por usuario
- Visualización de Tweets de cuentas relevantes

Requisitos de base de datos - Tweets

- Base de datos - Almacenamiento de nuevos Tweets
- Base de datos - Búsqueda de Tweets en la base de datos
- Base de datos - Visualización de Tweets almacenados
- Base de datos - Eliminación de Tweets almacenados

Requisitos de base de datos - Agrupaciones

- Base de datos - Búsqueda de agrupaciones de Tweets
- Base de datos - Visualización de agrupaciones de Tweets
- Base de datos - Edición de agrupaciones de Tweets
- Base de datos - Eliminación de agrupaciones de Tweets

Requisitos de base de datos - Cuentas relevantes

- Base de datos - Añadir nueva cuenta relevante
- Base de datos - Búsqueda de cuentas relevantes
- Base de datos - Visualización de cuentas relevantes
- Base de datos - Eliminación de cuentas relevantes

Requisitos de análisis

- Análisis - Vista de panel de análisis
- Análisis - Gráfica de difusión de Tweets
- Análisis - Tabla cuentas de mayor impacto
- Análisis - Tabla primeras cuentas de difusión
- Análisis - Mostrar retweets
- Análisis - Actualización de retweets
- Análisis - Obtención de tweets similares

Requisitos de agrupación de Tweets

- Agrupación - Agrupación de tweets
- Agrupación - Añadir Tweet a la agrupación
- Agrupación - Eliminación de tweet de agrupación

- Agrupación - Visualización de gráfica agregada

Requisitos de tracking

- Tracking - Activar trackeo automático
- Tracking - Desactivar trackeo automático
- Tracking - Periodicidad de actualización
- Tracking - Actualización de status

Requisitos de respuesta

- Respuesta - Crear una respuesta
- Respuesta - Validar respuesta
- Respuesta - Enviar respuesta

Requisitos de detección de Bots

- Detección de Bots

Requisitos no funcionales**Requisitos de interfaz**

- Estilo visual de Twitter
- Estilos de interfaz
- Diseño responsivo

Requisitos de seguridad

- Prevención de ataque por inyección SQL
- Prevención de ataque Cross Site Scripting
- Prevención de ataque de pérdida de autenticación
- Tratamiento de contraseñas
- Credenciales de uso de API Twitter

Requisitos de actualización

- Periodicidad de actualización

Requisitos de idiomas

- Registro de Idiomas

Otros requisitos no funcionales

- Registro de Logs

5.1.3. Definición de los requisitos del sistema

Requisitos de login y registro

Requisito	Registro de usuarios
Código	RF - 001
Prioridad	Esencial
Descripción	El sistema deberá permitir el registro de nuevos usuarios mediante un formulario que recoge un email válido, nombre de usuario, apellidos (opcional) y la contraseña (con doble verificación). El formulario debe verificar el email es válido y que la contraseña introducida contiene al menos cuatro dígitos.

Requisito	Login de usuario
Código	RF - 002
Prioridad	Esencial
Descripción	El sistema deberá obligar a los usuarios a iniciar sesión en el sistema para poder usar sus funcionalidades. El login debe hacerse mediante el correo y la contraseña de usuario.

Requisito	Recuperación de contraseña
Código	RF - 003
Prioridad	Esencial
Descripción	El sistema deberá permitir la recuperación de contraseña para usuarios no logados. Esta recuperación se hará vía solicitud desde la pantalla de login, mediante el envío de un email al usuario con un formulario para restablecer una nueva contraseña.

Requisito	Fin de sesión
Código	RF - 004
Prioridad	Esencial
Descripción	El sistema deberá permitir al usuario finalizar la sesión en cualquier momento, terminando su sesión en el sistema y redirigiendo al usuario a la pantalla de login.

Requisito	Visualización de perfil de usuario
Código	RF - 005
Prioridad	Esencial
Descripción	El sistema deberá permitir al usuario visitar su ficha de perfil de usuario con sus datos personales. La ficha debe mostrar los campos de correo, nombre y apellidos. El campo de contraseña no se mostrará en ningún caso.

Requisito	Actualización de perfil de usuario
Código	RF - 006
Prioridad	Esencial
Descripción	El sistema deberá permitir al usuario actualizar los campos de nombre y apellidos de su perfil de usuario a partir de su ficha de usuario.

Requisito	Eliminación de perfil de usuario
Código	RF - 007
Prioridad	Esencial
Descripción	El sistema deberá permitir al usuario su eliminación como usuario del sistema, lo que supondrá eliminar el registro correspondiente de la base de datos de usuario

Requisitos de obtención de Tweets

Requisito	Búsqueda de Tweets por tema
Código	RF - 008
Prioridad	Esencial
Descripción	El sistema deberá permitir la búsqueda de nuevos tweets a partir de un campo de búsqueda de la interfaz. A partir del/los términos introducidos por el usuario en el campo de búsqueda, el sistema tratará de recuperar hasta 15 tweets relacionados con esos términos.

Requisito	Búsqueda de Tweets por usuario
Código	RF - 009

Prioridad	Esencial
Descripción	El sistema deberá permitir la búsqueda de tweets a partir de un nombre de usuario de Twitter. Se tratará de recuperar al menos los 15 últimos tweets de la cuenta del usuario.
Requisito	Visualización de Tweets de cuentas relevantes
Código	RF - 010
Prioridad	Deseable
Descripción	El sistema mostrará al usuario los últimos 3 Tweets que han producido cada una de las cuentas almacenadas en la lista de cuentas relevantes

Requisitos de base de datos

Requisito	Almacenamiento de nuevos Tweets
Código	RF - 011
Prioridad	Esencial
Descripción	El sistema deberá permitir el almacenamiento en la base de datos de los tweets obtenidos por usuario, tema o por seguimiento de cuentas relevantes

Requisito	Búsqueda de Tweets en la base de datos
Código	RF - 012
Prioridad	Esencial
Descripción	El sistema deberá permitir al usuario la búsqueda de tweets almacenados en la base de datos.

Requisito	Visualización de Tweets almacenados
Código	RF - 013
Prioridad	Esencial
Descripción	El sistema deberá permitir al usuario la visualización de la información que contiene la ficha de cualquiera de los tweets almacenados en la base de datos

Requisito	Eliminación de Tweets almacenados
Código	RF - 014
Prioridad	Esencial

Descripción	El sistema deberá permitir al usuario la eliminación de cualquiera de los tweets que se encuentran almacenados en la base de datos, eliminando de esta forma cualquier tipo de información relacionada con el mismo
--------------------	---

Base de datos - Agrupaciones

Requisito	Búsqueda de agrupaciones de Tweets
Código	RF - 015
Prioridad	Esencial
Descripción	El sistema deberá permitir al usuario la búsqueda en la base de datos de las diferentes agrupaciones de tweets creadas.

Requisito	Visualización de agrupaciones de Tweets
Código	RF - 016
Prioridad	Esencial
Descripción	El sistema deberá permitir al usuario la visualización de información almacenada en la base de datos de cualquier agrupación de tweets creada

Requisito	Edición de agrupaciones de Tweets
Código	RF - 017
Prioridad	Esencial
Descripción	El sistema deberá permitir al usuario la edición de los campos almacenados en la base de datos con respecto a cada una de las agrupaciones de tweets de la base de datos

Requisito	Eliminación de agrupaciones de Tweets
Código	RF - 018
Prioridad	Esencial
Descripción	El sistema deberá permitir al usuario la eliminación de la base de datos de las agrupaciones de tweets almacenadas

Base de datos – Cuentas relevantes

Requisito	Añadir cuenta relevante
Código	RF - 019

Prioridad	Deseable
Descripción	El sistema deberá permitir al usuario la opción de añadir cuentas de usuario a la lista de cuentas relevantes
Requisito	Búsqueda de cuentas relevantes
Código	RF - 020
Prioridad	Deseable
Descripción	El sistema deberá permitir al usuario la búsqueda y recuperación de las cuentas relevantes almacenadas en la base de datos
Requisito	Visualización de cuenta relevante
Código	RF - 021
Prioridad	Deseable
Descripción	El sistema deberá permitir al usuario la opción de visualizar la información de la ficha de "cuenta relevante" almacenada en la base de datos
Requisito	Eliminación de cuenta relevante
Código	RF - 022
Prioridad	Deseable
Descripción	El sistema deberá permitir al usuario la opción de eliminar de la base de datos cualquiera de las cuentas relevantes almacenadas

Requisitos de análisis

Requisito	Vista del panel de análisis
Código	RF - 023
Prioridad	Esencial
Descripción	El sistema deberá proporcionar al usuario una vista en la que se muestre el tweet que se desea analizar, junto con las opciones que permitan al usuario realizar las acciones de actualizar retweets, análisis de tweet, eliminar el tweet, la búsqueda de tweets similares, enviar una respuesta al tweet y activar la opción de tracking de tweet

Requisito	Gráfica de difusión de Tweets
Código	RF - 024
Prioridad	Esencial
Descripción	El panel de análisis deberá proporcionar al usuario una gráfica en la que se muestre la difusión de un tweet a lo largo del tiempo, a partir de los datos de su fecha de creación y de las fechas de los sucesivos retweets.

Requisito	Tabla de cuentas de mayor impacto
Código	RF - 025
Prioridad	Esencial
Descripción	<p>El panel de análisis deberá mostrar al usuario una tabla con información relevante de las diez cuentas de mayor impacto a la hora de difundir un tweet ordenadas de mayor a menor impacto. Se considera que una cuenta tiene más impacto sobre otra si tiene mayor número de seguidores. Esta tabla engloba el tanto el análisis de un solo tweet como el de una agrupación de tweets.</p> <p>Se considera información relevante la que contenga al menos el nombre de usuario y su número de seguidores.</p>

Requisito	Tabla de primeras cuentas de difusión
Código	RF - 026
Prioridad	Esencial
Descripción	El panel de análisis deberá mostrar al usuario una tabla con información relevante de las diez primeras cuentas encargadas de difundir un tweet o una noticia según la fecha y hora de difusión. Esta tabla engloba el tanto el análisis de un solo tweet como el de una agrupación de tweets. Se considera información relevante la que contenga al menos el nombre de usuario y su número de seguidores.

Requisito	Mostrar retweets
Código	RF - 027
Prioridad	Esencial
Descripción	El panel de análisis deberá mostrar el número de retweets tanto del tweet analizado como el de cada uno de los posibles tweets que componen una agrupación de tweets.

Requisito	Actualización de retweets
Código	RF - 028
Prioridad	Esencial
Descripción	El sistema deberá proporcionar al usuario la opción de actualizar los retweets de un tweet o de cada uno de los tweets que componen una agrupación de tweets.

Requisito	Obtención de Tweets similares
Código	RF - 029
Prioridad	Esencial
Descripción	El panel de análisis deberá proporcionar al usuario la opción de obtener tweets similares al que se está analizando. Se consideran tweets similares aquellos que contienen una url con la misma noticia. En caso de que el tweet inicial no cuente con una url, se mostrará un mensaje de aviso y no se realizará ninguna acción.

Requisitos de agrupación de Tweets

Requisito	Agrupación de Tweets
Código	RF - 030
Prioridad	Esencial
Descripción	El sistema deberá proporcionar al usuario la opción de crear una agrupación de tweets a partir del tweet analizado y cualquiera de los tweets similares recuperados.

Requisito	Añadir Tweet a la agrupación
Código	RF - 032
Prioridad	Esencial
Descripción	El panel de análisis deberá permitir al usuario añadir a la agrupación de tweets cualquiera de los tweets similares obtenidos mediante la búsqueda.

Requisito	Eliminación de Tweet de la agrupación
Código	RF - 031
Prioridad	Esencial

Descripción	El panel de análisis deberá permitir al usuario eliminar un tweet de una agrupación de tweets creada anteriormente. Como mínimo la agrupación deberá contener el tweet principal que se está analizando.
Requisito	Visualización de gráfica agregada
Código	RF - 033
Prioridad	Esencial
Descripción	El panel de análisis deberá mostrar al usuario una gráfica agregada formada por la unión de cada uno de los tweets que componen una agrupación, así como de todos sus retweets. En la gráfica se mostrará la difusión de los tweets y sus retweets a lo largo del tiempo.

Requisitos de tracking

Requisito	Activar trackeo automático
Código	RF - 034
Prioridad	Esencial
Descripción	El sistema deberá permitir al usuario activar la opción de trackeo de tweets o de agrupaciones de tweets. El trackeo de tweets consistirá en un estado en el que el número de retweets y el estado del tweet se actualiza de forma periódica. La periodicidad de actualización debe ser de al menos una vez diaria y realizarse se acuerdo a los criterios establecidos en los requisitos no funcionales.
Requisito	Desactivar el trackeo automático
Código	RF - 035
Prioridad	Esencial
Descripción	El sistema deberá permitir al usuario desactivar en cualquier momento la opción de trackeo automático, dejando así de actualizar su estatus hasta que vuelva a ser activado.
Requisito	Actualización de estatus
Código	RF - 036
Prioridad	Esencial

Descripción	<p>Si el sistema detecta en algún momento que el estatus de un tweet ha cambiado en la red social a borrado o privado, el sistema eliminará el tweet de la base de datos del sistema, así como la información relacionada al mismo que haya podido ser almacenada.</p> <p>Este requisito se mantendrá en vigor para cumplir con la política de privacidad de Twitter, y se mantendrá siempre actualizado en función de la misma.</p>
--------------------	--

Requisitos de respuesta

Requisito	Crear una respuesta
Código	RF - 037
Prioridad	Esencial
Descripción	El sistema deberá proporcionar al usuario a través del panel de análisis la opción de escribir o editar una respuesta para el tweet o la agrupación de tweets que se está analizando.

Requisito	Validar una respuesta
Código	RF - 038
Prioridad	Esencial
Descripción	El sistema deberá validar que la respuesta escrita por el usuario cumpla con el límite de caracteres establecido por Twitter en su política actual.

Requisito	Enviar una respuesta
Código	RF - 039
Prioridad	Esencial
Descripción	El sistema deberá permitir el envío de la respuesta creada por el usuario y que será enviada a Twitter como respuesta al tweet analizado o a cada uno de los tweets que componen una agrupación de tweets

Requisitos de detección de bots

Requisito	Detección de Bots
Código	RF - 040
Prioridad	Deseable

Descripción	El sistema deberá proporcionar al usuario la opción de analizar la probabilidad de que una cuenta concreta sea un Bot. Para ello el sistema puede hacer uso de otros sistemas externos como Botometer.
--------------------	--

Requisitos no funcionales

Requisitos de interfaz

Requisito	Estilo visual de Twitter
Código	RNF - 001
Prioridad	Deseable
Descripción	El sistema deberá proporcionar una interfaz que permita al usuario reconocer la relación del sistema con Twitter. Los tweets mostrados a través de la interfaz deberán presentar el nombre de usuario, fecha y texto en un formato similar al de Twitter. Se utilizará para ello la gama de colores de Twitter disponible en el apartado de diseño del sistema.

Requisito	Estilos de interfaz
Código	RNF - 002
Prioridad	Opcional
Descripción	El sistema deberá proporcionar una interfaz que tenga una apariencia seria y elegante, evitando las gamas de colores llamativas para permitir al usuario centrar su atención la funcionalidad de la herramienta. Se usarán una gama de colores de negros, grises, blancos y azules.

Requisito	Diseño responsivo
Código	RNF - 022
Prioridad	Opcional
Descripción	La interfaz del sistema deberá adoptar un diseño responsive, para adaptarse a la vista tanto en dispositivos móviles como en tablets, considerándose tamaños de ancho de pantalla de al menos 320px.

Requisitos de seguridad

Requisito	Prevención de ataque por inyección SQL
Código	RNF - 003
Prioridad	Opcional
Descripción	El sistema deberá prevenir ataques por inyección SQL en la base de datos. Para ello usarán estándares validado por organizaciones especializadas.

Requisito	Prevención de ataque Cross Site Scripting
Código	RNF - 004
Prioridad	Esencial
Descripción	El sistema deberá prevenir ataques Cross Site tanto directo como indirecto. Para ello usará un sistema construido y validado por organizaciones especializadas

Requisito	Prevención de ataque de pérdida de autenticación
Código	RNF - 005
Prioridad	Esencial
Descripción	El sistema deberá prevenir ataques de pérdida de autenticación de usuarios. Para ello usará un sistema construido y validado por organizaciones especializadas

Requisito	Tratamiento de contraseñas
Código	RNF - 006
Prioridad	Esencial
Descripción	El sistema deberá tratar las contraseñas con un cifrado SHA512. En ningún momento el sistema podrá mostrar las contraseñas almacenadas a ningún usuario, incluyendo al propietario de las mismas.

Requisito	Credenciales de uso de API Twitter
Código	RNF - 007
Prioridad	Esencial
Descripción	El sistema deberá mantener de forma segura las credenciales para el uso de la API de Twitter en un archivo externo y fuera del código.

Requisitos de actualización

Requisito	Periodicidad de actualización
Código	RNF - 008
Prioridad	Deseable
Descripción	Cuando un tweet se encuentra en estado de tracking, debe actualizarse de forma automática al menos una vez al día. Esta actualización deberá hacerse en los periodos de menor actividad, en segundo plano o en periodos nocturnos de la franja horaria de los usuarios del sistema (franja nocturna 3:00 am a 6:00 am).

Requisito	Registro de Logs
Código	RNF - 009
Prioridad	Deseable
Descripción	El sistema deberá producir un registro de logs con los accesos y los errores del sistema en un archivo del servidor.

Requisitos de idiomas

Requisito	Registro de Idiomas
Código	RNF - 010
Prioridad	Esencial
Descripción	El sistema deberá estar disponible en lengua inglesa, siendo opcional la inclusión de otras lenguas

5.2. Especificación de casos de uso

5.2.1. Diagrama de casos de uso

El diagrama de casos de uso tiene principalmente dos objetivos: el primero para reflejar la relación del sistema con los diferentes actores como pueden ser los usuarios u otros sistemas, y en segundo lugar, para permitir un desarrollo correcto del sistema, englobando grupos de requisitos que guarden relación y permitan seguir un desarrollo por fases.

Si bien el diagrama de casos de uso refleja una serie de acciones y de interacciones que un usuario o sistema externo realiza con el propio sistema, cada caso de uso puede englobar en sí mismo más de una posible acción o diferentes requisitos. Estas agrupaciones quedarán reflejadas en la matriz de referencias cruzadas y en los casos de uso en formato expandido.

NOTA:

Es importante mencionar, que, aunque algunos casos de uso son más sencillos o atómicos que otros, se ha decidido separarlo de otros si implica a otros actores, si no guarda especial relación con otros casos de uso. Además, cuando un caso de uso engloba un alto número de requisitos, se ha separado en dos casos de uso para facilitar las fases de desarrollo. Por último, algunos casos de uso se han separado de otros aunque engloben un bajo número de requisitos cuando éstos conllevan una dificultad mayor de lo normal a la hora de su desarrollo (como es el caso del caso de uso de obtención de tweets similares)

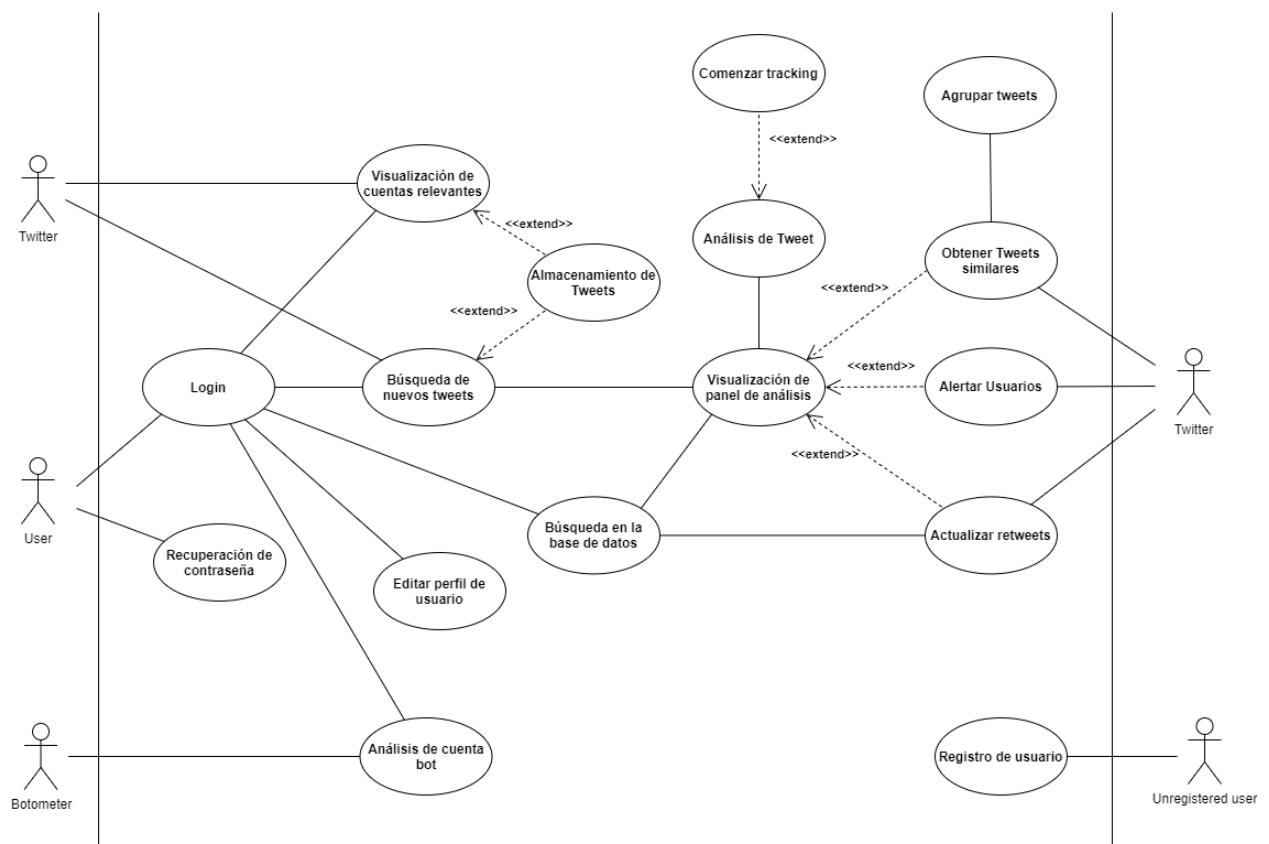


Ilustración 10. Diagrama de casos de uso

Para la creación de los casos de uso se ha usado la documentación ofrecida en la especificación UML 2.5 (Unified Modeling Language). Esta especificación es un estándar usado en la ingeniería del software, que permite la estandarización del diseño del software. La documentación UML recoge los diagramas de casos de uso como un tipo de diagrama de comportamiento del sistema en el que se muestran los diferentes actores que interactúan con el mismo y las diferentes acciones que pueden realizar o que suponen cambios relevantes en el estado del sistema.

5.2.2. Justificación de la plantilla de casos de uso

Para la exposición detallada de cada uno de los casos de uso, se va a utilizar la plantilla que se muestra a continuación, que contiene los elementos que definen cada caso de uso:

Caso de uso:	Nombre del caso de uso
<i>Identificador:</i>	Código de identificación alfanumérico secuencial: CU-XXX
<i>Actores</i>	Actores implicados en el caso de uso
<i>Descripción</i>	Descripción ampliada del caso de uso
<i>Prioridad</i>	Esencial / Deseable
<i>Referencias cruzadas</i>	Dependencias y relaciones con otros casos de uso

5.3. Casos de uso en formato expandido

A continuación, se exponen todos los casos de uso en formato expandido, donde se recogen los casos de uso, así como su curso de ventos normal y cada uno de los cursos de eventos alternativos mostrando las acciones de los actores y la respuesta del sistema ante cada una de ellas.

Caso de uso	Registro de nuevo usuario
<i>Identificador:</i>	CU - 001
<i>Actores</i>	Usuario no registrado
<i>Descripción</i>	Un usuario que no estaba registrado en el sistema utiliza el formulario de registro para crear una nueva cuenta. Para ello introduce el nombre, apellido (opcional), un email y una contraseña.
<i>Prioridad</i>	Primario, esencial

Referencias cruzadas

-

Curso de eventos normal

Acción de los actores	Respuesta del sistema
1. El usuario no registrado, activa la opción de registro de usuario	
	2. El sistema muestra un formulario de registro con los siguientes campos: <ul style="list-style-type: none"> - Nombre - Apellidos (Opcional) - Email - Contraseña - Confirmación de contraseña
3. El usuario, rellena cada uno de los campos del formulario	
	4. El sistema, valida que todos los campos del formulario son correctos
	5. El introduce al usuario en la base de datos y registra el login de usuario
	6. El sistema redirige al usuario a la pantalla de búsqueda de Tweets

Curso de ventos alternativo

a) Error de formulario

Acción de los actores	Respuesta del sistema
3. El usuario, rellena cada uno de los campos del formulario	
	4. El sistema, valida que todos los campos del formulario son correctos
	5. El formulario no pasa la fase de validación por encontrar algún campo vacío o incorrecto

	6. El sistema muestra al usuario el tipo de error encontrado en el formulario
--	---

b) Error de la base de datos

Acción de los actores	Respuesta del sistema
	4. El sistema, valida que todos los campos del formulario son correctos
	5. El sistema trata de almacenar al usuario en la base de datos, pero se produce un error.
	6. El sistema informa al usuario del error, y no se produce el registro del usuario.

Caso de uso	Login de usuario
<i>Identificador:</i>	CU - 002
<i>Actores</i>	Usuario
<i>Descripción</i>	Un usuario registrado y no logado entra en el sistema a través de la pantalla de login usando sus credenciales de acceso (email y contraseña)
<i>Prioridad</i>	Primario, esencial
<i>Referencias cruzadas</i>	CU – 001, El usuario debe estar registrado

Curso de eventos normal

Acción de los actores	Respuesta del sistema
1. El usuario entra en la pantalla de login del sistema	
2. El usuario introduce las credenciales de acceso	
	3. El módulo de Control de Accesos del sistema valida las credenciales y registra al usuario como logado
	4. El sistema redirige al usuario a la pantalla de búsqueda de Tweets

Curso de eventos alternativo

a) Error de login

Acción de los actores	Respuesta del sistema
	3. El módulo de Control de Accesos detecta un error en las credenciales y bloquea el acceso del usuario
	4. El sistema alerta al usuario del error de acceso

Caso de uso	Recuperación de contraseña
<i>Identificador:</i>	CU - 003
<i>Actores</i>	Usuario
<i>Descripción</i>	Un usuario registrado en el sistema usa la opción de solicitar la recuperación de contraseña a partir de su email de registro
<i>Prioridad</i>	Secundario, deseado
<i>Referencias cruzadas</i>	CU – 001, El usuario debe estar registrado

Curso de eventos normal

Acción de los actores	Respuesta del sistema
1. El usuario activa la opción de recuperación de contraseña del menú de login	
2. El usuario introduce el email y activa la opción Solicitar nueva contraseña	
	3. El sistema valida el campo de email del formulario
	4. El sistema recupera el email de la base de datos
	5. El sistema envía un email al usuario hacia el formulario de la nueva contraseña
	6. El sistema alerta al usuario de que se ha enviado un email

Curso de eventos alternativo

a) Error de formulario

Acción de los actores	Respuesta del sistema
2. El usuario introduce el email y activa la opción Solicitar nueva contraseña	
	3. El formulario no pasa el proceso de validación, el sistema alerta al usuario

b) Usuario no reconocido

Acción de los actores	Respuesta del sistema
2. El usuario introduce el email y activa la opción Solicitar nueva contraseña	
	3. El sistema valida el formulario
	4. El sistema no encuentra al usuario en la base de datos
	5. El sistema muestra un mensaje de error al usuario indicando que el usuario no existe

Caso de uso	Editar perfil de usuario
<i>Identificador:</i>	CU - 004
<i>Actores</i>	Usuario registrado
<i>Descripción</i>	Un usuario registrado en el sistema accede a la visualización de su perfil de usuario y edita el nombre y los apellidos
<i>Prioridad</i>	Secundaria, deseado
<i>Referencias cruzadas</i>	CU – 001, El usuario debe estar registrado

Curso de eventos normal

Acción de los actores	Respuesta del sistema
1. El usuario accede al sistema a través del login	
2. El usuario accede a su perfil de usuario	
3. El usuario edita el nombre y los apellidos	
	4. El sistema valida el formulario y registra los cambios en la base de datos
	5. El sistema informa al usuario del éxito de la operación

Curso de eventos alternativo

a) Error de base de datos

Acción de los actores	Respuesta del sistema
2. El usuario edita el nombre y los apellidos	
	3. El sistema valida el formulario y se produce un error al almacenar los cambios en la base de datos
	4. Se envía una alerta al usuario informando del error

Caso de uso:	Búsqueda de nuevos tweets
Identificador:	CU - 005
Actores	Usuario, Twitter
Descripción	El usuario utiliza el panel de búsqueda del sistema para la búsqueda y recuperación de tweets a partir de palabras clave, noticias o nombres de usuario.
Prioridad	Primario, esencial
Referencias cruzadas	CU – 002, el usuario debe estar identificado en el sistema

Curso de eventos normal

Acción de los actores	Respuesta del sistema

1. El usuario usa el menú para acceder al panel de búsqueda de Tweets	
2. El usuario escribe una palabra, título de noticia o url en la barra de búsqueda	
3. El usuario activa la opción de buscar	
	4. El sistema recoge la petición, y realiza una llamada a la API de Twitter de búsqueda de tweets
5. El actor Twitter, procesa la petición a través de la API, recupera los tweets y los envía al sistema	
	6. El sistema recoge la respuesta de Twitter, y la procesa para devolvérsela al usuario
	7. Se envía al usuario los tweets obtenidos de Twitter
	8. Se muestran los tweets a través de la interfaz de usuario
9. El usuario visualiza los Tweets obtenidos, haciendo scroll a través de la pantalla	

Curso de eventos alternativo

a) Búsqueda por usuario

Acción de los actores	Respuesta del sistema
1.El usuario usa el menú para acceder al panel de búsqueda de Tweets	
2. El usuario activa la opción de “búsqueda por usuario” en la barra de búsqueda	
3. El usuario introduce la cuenta de usuario que quiere buscar	
4. El usuario activa la opción de buscar	
	5. El sistema recoge la petición, y realiza una llamada a la API de Twitter de búsqueda de tweets

b) El campo de búsqueda está vacío

Acción de los actores	Respuesta del sistema
2. El usuario pulsa el botón de búsqueda sin introducir nada en la barra de búsqueda	
	3. El sistema muestra una alerta, indicando que el campo de búsqueda está vacío

c) No se encuentran tweets relacionados

Acción de los actores	Respuesta del sistema
	6. El sistema recoge la respuesta de Twitter, que no contiene tweets
	7. El sistema muestra una alerta en la vista, indicando al usuario el tipo de error

Caso de uso	Almacenamiento de tweets
<i>Identificador:</i>	CU - 006
<i>Actores</i>	Usuario
<i>Descripción</i>	El usuario, ya sea mediante la búsqueda de un tweet a través del panel de búsqueda o mediante el panel de cuentas relevantes, almacena en la base de datos cualquiera de los tweets para su posterior análisis
<i>Prioridad</i>	Primario, esencial
<i>Referencias cruzadas</i>	CU – 005, CU – 007, se debe haber activado la búsqueda de tweets a través del panel de búsqueda o del feed de cuentas relevantes

Curso de eventos normal

Acción de los actores	Respuesta del sistema
1. El usuario, a partir de la búsqueda de nuevos tweets ha obtenido resultados de búsqueda	
2. El usuario detecta un tweet que quiere almacenar y pulsa el botón de guardar	

	3. El sistema almacena en la base de datos el tweet indicado por el usuario
	4. El sistema redirige de forma automática al usuario al panel de búsqueda

Curso de eventos alternativo**a) Acceso desde el panel de cuentas relevantes**

Acción de los actores	Respuesta del sistema
1. El usuario visualiza los últimos tweets de cuentas relevantes	
2. El usuario detecta un tweet que quiere almacenar y pulsa el botón de guardar	

b) Error al almacenar

Acción de los actores	Respuesta del sistema
	3. Se produce un error al almacenar el tweet en la base de datos
	4. El sistema muestra una alerta al usuario indicando el error

Caso de uso	Visualización de cuentas relevantes
<i>Identificador:</i>	CU - 007
<i>Actores</i>	Usuario, Twitter
<i>Descripción</i>	El usuario a través del panel de cuentas relevantes accede a los últimos tweets que provienen de cuentas marcadas como relevantes por el usuario
<i>Prioridad</i>	Secundario, deseado
<i>Referencias cruzadas</i>	-

Curso de eventos normal

Acción de los actores	Respuesta del sistema
1.El usuario usa el menú para acceder al panel de cuentas relevantes	
	2. El sistema hace una llamada a la API de Twitter para obtener los últimos tweets de las cuentas que sigue el usuario
3. El Actor Twitter, procesa la petición y devuelve al sistema el feed de tweets de las cuentas que sigue el usuario	
	4. El sistema procesa la respuesta y se la devuelve a la vista
	5. La interfaz muestra los tweets obtenidos
6. El usuario visualiza los Tweets obtenidos, haciendo scroll a través de la pantalla	

Curso de eventos alternativo

a) El usuario no sigue a ninguna cuenta

Acción de los actores	Respuesta del sistema
3. El Actor Twitter, procesa la petición y no detecta cuentas. Devuelve una lista vacía	
	4. El sistema procesa la respuesta y no detecta tweets
	5. EL sistema muestra un mensaje al usuario de alerta indicando que no se han detectado cuentas relevantes

Caso de uso	Búsqueda en la base de datos
Identificador:	CU - 008
Actores	Usuario, Twitter

<i>Descripción</i>	El usuario, a través del menú accede a la visualización de la base de datos para buscar un Tweet o una agrupación de Tweets almacenada
<i>Prioridad</i>	Primario, esencial
<i>Referencias cruzadas</i>	CU – 006, existen tweets almacenados en la base de datos

Curso de eventos normal

Acción de los actores	Respuesta del sistema
1. El usuario accede a través del menú a la vista de la base de datos	
	2. El sistema muestra en formato de tabla todos los tweets almacenados en la base de datos junto con las opciones de visualizar, editar, actualizar, trackear y eliminar.
3. El usuario hace una búsqueda en la barra del buscador de la base de datos para filtrar los resultados	
	4. El sistema filtra los resultados de acuerdo al buscador y los muestra al usuario
5. El usuario encuentra el Tweet que estaba buscando y accede a su ficha	
	6. El sistema muestra al usuario todos los campos almacenados de ese Tweet

Curso de eventos alternativo

a) Acceso a una agrupación de Tweets

Acción de los actores	Respuesta del sistema
1. El usuario accede a través del menú a la vista de la base de datos de agrupación de Tweets	
	2. El sistema muestra en formato de tabla todas las <i>agrupaciones de Tweets</i> almacenadas en la base de

	datos junto con las opciones de visualizar, editar, actualizar, trackear y eliminar.
3. El usuario hace una búsqueda en la barra del buscador de la base de datos para filtrar los resultados	
	4. El sistema filtra los resultados de acuerdo al buscador y los muestra al usuario
5. El usuario encuentra la agrupación de Tweets que estaba buscando y accede a su ficha	
	6. El sistema muestra al usuario todos los campos almacenados de esa agrupación

b) Eliminación de un registro

Acción de los actores	Respuesta del sistema
	2. El sistema muestra en formato de tabla todas las agrupaciones de Tweets almacenadas en la base de datos junto con las opciones de visualizar, editar, actualizar, trackear y eliminar.
3. El usuario utiliza la opción de eliminar registro de la base de datos	
	4. El sistema elimina el registro de la base de datos y todas las vinculaciones que existan en cascada
	5. Se informa al usuario del resultado de la operación

Caso de uso	Visualización de panel de análisis
Identificador:	CU - 009
Actores	Usuario
Descripción	<p>El usuario, a través del panel de análisis visualiza el tweet que es objetivo de análisis, junto con un menú de opciones que contiene las acciones que puede realizar.</p> <p>Dichas opciones son:</p> <ul style="list-style-type: none"> - Analizar - Actualizar retweets - Buscar tweets similares

	<ul style="list-style-type: none"> - Alertar a usuario - Comenzar tracking - Eliminar tweet
<i>Prioridad</i>	Primario, esencial
<i>Referencias cruzadas</i>	CU – 005, CU – 006, el usuario a partir de un tweet almacenado o de la búsqueda de tweets procede a analizarlos

Curso de eventos normal

Acción de los actores	Respuesta del sistema
1. El usuario accede al panel de análisis a través del menú	
	2. El sistema muestra al usuario el tweet objeto de análisis, así como el submenú de acciones disponibles para el análisis

Curso de eventos alternativo**a) Acceso desde la base de datos de tweets**

Acción de los actores	Respuesta del sistema
1. El usuario accede al panel de análisis a través de la tabla por la que se visualiza la base de datos de tweets almacenados	
	2. El sistema muestra al usuario el tweet objeto de análisis, así como el submenú de acciones disponibles para el análisis

b) Acceso desde la base de datos de agrupaciones de tweets

Acción de los actores	Respuesta del sistema
1. El usuario accede al panel de análisis a través de la tabla por la que se visualiza la base de datos de agrupaciones de tweets	

	2. El sistema muestra al usuario el tweet objeto de análisis, así como el submenú de acciones disponibles para el análisis
	3. El sistema muestra al usuario en una tabla el resto de los tweets que forman parte de la agrupación

Caso de uso	Búsqueda de tweets similares
<i>Identificador:</i>	CU – 0010
<i>Actores</i>	Usuario, Twitter
<i>Descripción</i>	El usuario, a través del panel de análisis activa la búsqueda de tweets similares, obteniendo nuevos tweets que el sistema detecta como similares al que se está analizando
<i>Prioridad</i>	Secundario, deseado
<i>Referencias cruzadas</i>	CU – 009, la búsqueda de tweets similares se hace en relación al tweet que se está analizando

Curso de eventos normal

Acción de los actores	Respuesta del sistema
1. El usuario, desde el panel de análisis activa la opción de búsqueda de tweets similares desde el panel de análisis	
	2. El sistema recoge la url que se encuentra embebida en el tweet
	3. El sistema realiza una petición http/https a la web de la url, y recupera la página a la que pertenece
	4. Mediante técnicas de web scrapping, se recoge el título de la noticia
	5. Se realiza una llamada a la API de Twitter con el título obtenido
6. El actor Twitter procesa la petición y recupera los tweets relacionados con la noticia	
	7. El sistema procesa los tweets recuperados, y los muestra al usuario en el panel de análisis

Curso de eventos alternativo

a) El tweet no contiene ninguna url

Acción de los actores	Respuesta del sistema
1. El usuario, desde el panel de análisis activa la opción de búsqueda de tweets similares desde el panel de análisis	
	2. El sistema no detecta ninguna url contenida en el tweet
	3. El sistema devuelve un mensaje de error indicando que no ha detectado ninguna noticia

b) El tweet no puede recuperar la página que contiene la url

Acción de los actores	Respuesta del sistema
	3. El sistema realiza una petición http/https a la web de la url, y recupera la página a la que pertenece
	4. El sistema no es capaz de recuperar la página o de detectar el título de la noticia
	5. El sistema muestra una alerta al usuario indicando el error

c) No se encuentran tweets similares

Acción de los actores	Respuesta del sistema
	5. Se realiza una llamada a la API de Twitter con el título obtenido
6. El actor Twitter procesa la petición y no encuentra ningún tweet relacionado con la petición	
	7. El sistema procesa la respuesta, detecta que no se han recuperado tweets y muestra una alerta al usuario

Caso de uso	Agrupar tweets
<i>Identificador:</i>	CU – 0011
<i>Actores</i>	Usuario
<i>Descripción</i>	El usuario a través del panel de análisis crea agrupaciones de tweets que considera que son similares para su análisis de forma conjunta
<i>Prioridad</i>	Secundario, deseado
<i>Referencias cruzadas</i>	CU – 0010, el usuario debe haber realizado una búsqueda de tweets similares previamente a agruparlos

Curso de eventos normal

Acción de los actores	Respuesta del sistema
1. El usuario tras haber realizado una búsqueda de tweets similares, se encuentra en el panel de búsqueda con los tweets similares recuperados por el sistema	
2. El usuario marca aquellos tweets que considera similares activando el botón de añadir	
	3. El sistema incluye los tweets marcados por el usuario en una tabla en la que se visualizan los cambios
	4. El sistema actualiza la base de datos y guarda los tweets incluidos como una agrupación

Curso de eventos alternativo

a) Acceso desde la base de datos

Acción de los actores	Respuesta del sistema
1. El usuario, tras seleccionar un tweet o una agrupación de tweets desde la base de datos es redirigido al panel de análisis	
	2. El sistema, detecta que el usuario ha sido redirigido desde la vista de tabla de la base de datos, y realiza una búsqueda en la base de datos

	3. El sistema recupera los tweets que forman parte de la agrupación
	4. El sistema incluye los tweets marcados por el usuario en una tabla en la que se visualizan los tweets que forman parte de la agrupación

b) Eliminación de un tweet de la agrupación

Acción de los actores	Respuesta del sistema
	4. El sistema incluye los tweets marcados por el usuario en una tabla en la que se visualizan los tweets que forman parte de la agrupación
5. El usuario, elimina uno de los tweets del grupo a través de la tabla de visualización de la agrupación	
	6. El sistema detecta la acción del usuario y extrae el tweet de la tabla que contiene la agrupación de tweets
	7. El sistema elimina de la base de datos el tweet mencionado, que deja de formar parte de la agrupación

c) No se encuentran tweets relacionados

Acción de los actores	Respuesta del sistema
	2. El sistema, detecta que el usuario ha sido redirigido desde la vista de tabla de la base de datos, y realiza una búsqueda en la base de datos
	3. No se detectan tweets relacionados con la agrupación, se muestra un mensaje de alerta al usuario

d) Error en la búsqueda de la base de datos

Acción de los actores	Respuesta del sistema

	2. El sistema, detecta que el usuario ha sido redirigido desde la vista de tabla de la base de datos, y realiza una búsqueda en la base de datos
	3. Se produce un error de lectura o escritura en la base de datos
	4. El sistema muestra una alerta al usuario mostrando el error

Caso de uso	Analizar Tweet
<i>Identificador:</i>	CU - 012
<i>Actores</i>	Usuario
<i>Descripción</i>	<p>El usuario tiene opción de visualizar un análisis de un tweet o de un grupo de tweets a través del panel de análisis.</p> <p>Como elementos de la visualización se incluyen:</p> <ul style="list-style-type: none"> - Gráficas de difusión del tweet en el tiempo - Gráfica agregada de difusión de una agrupación de tweets - Tabla "Top fake users" Con los usuarios implicados en la difusión ordenado por número de seguidores - Tabla "Top first users" Tabla con los usuarios implicados en la difusión del tweet ordenados por fecha - Alcance potencial: Con el número agregado de seguidores de cada uno de los usuarios implicados en la difusión - Número total de retweets alcanzado entre todos los usuarios implicados en la difusión
<i>Prioridad</i>	Primario, esencial
<i>Referencias cruzadas</i>	CU – 009, el análisis se hace desde el panel de análisis

Curso de eventos normal

Acción de los actores	Respuesta del sistema
1. El usuario accede al panel de análisis redirigido tras haber almacenado un tweet en el caso de uso Almacenar Tweets	
2. El usuario acciona la opción de analizar Tweet	
	3. El sistema recupera de la base de datos todos los retweets del tweet que se está analizando

	4. El sistema muestra una gráfica temporal en la que se visualiza cómo se ha difundido el tweet a lo largo del tiempo y cada retweet.
	5. El sistema muestra una tabla con los usuarios más relevantes a la hora de difundir el tweet, compuestos por los usuarios que han retuiteado ordenados por número de seguidores
	6. El sistema muestra el alcance potencial del tweet (La suma del número de seguidores del tweet original y de cada uno de los usuarios que ha retuiteado)
	7. El sistema muestra una tabla con los usuarios relevantes a la hora de difundir el tweet ordenados por fecha de interacción
8. El usuario interactúa con los gráficos y los diferentes elementos de la pantalla	

Curso de eventos alternativo

a) El usuario accede desde el menú

Acción de los actores	Respuesta del sistema
1. El usuario accede al panel de análisis a través del menú	
	2. El sistema recupera de la base de datos el último tweet almacenado disponible

EL usuario accede a través del menú y no hay tweets almacenados

Acción de los actores	Respuesta del sistema
1. El usuario accede al panel de análisis a través del menú	
	2. El sistema recupera de la base de datos el último tweet almacenado disponible
	3. El sistema no detecta tweets almacenados, por lo que genera una alerta al usuario

b) El usuario accede a través del panel de cuentas relevantes

Acción de los actores	Respuesta del sistema
1. El usuario accede al panel de análisis redirigido al almacenar un tweet desde el panel de cuentas relevantes	
	2. El sistema muestra en el panel de análisis el tweet escogido por el usuario junto con las opciones de análisis

Caso de uso	Actualizar retweets
<i>Identificador:</i>	CU - 013
<i>Actores</i>	Usuario, Twitter
<i>Descripción</i>	El usuario desde el panel de análisis o desde la vista de tablas de la base de datos, actualizar retweets del tweet analizado. El sistema muestra los nuevos retweets y actualiza las gráficas y los datos
<i>Prioridad</i>	Primaria, esencial
<i>Referencias cruzadas</i>	CU – 009, los retweets se actualizan desde el panel de análisis

Curso de eventos normal

Acción de los actores	Respuesta del sistema
1. El usuario, desde el panel de análisis, activa la opción de actualizar retweets	
	2. El sistema recoge el id del tweet que se quiere actualizar, y realiza una llamada a la API de Twitter para obtener los retweets

3. El actor Twitter, procesa la petición y devuelve los retweets relativos a la petición	
	4. El sistema procesa la respuesta, y almacena en la base de datos los nuevos retweets obtenidos
	5. El sistema actualiza los datos del panel de análisis y las correspondientes gráficas

Curso de eventos alternativo

a) Error en la API de Twitter

Acción de los actores	Respuesta del sistema
3. El actor Twitter, procesa la petición y se produce un error	
	4. El sistema detecta el error, y alerta al usuario con el tipo de error obtenido de Twitter

b) Actualización de agrupación de tweets

Acción de los actores	Respuesta del sistema
1. El usuario, desde el panel de análisis, activa la opción de actualizar retweets	
	2. El sistema recoge el id de cada uno de los tweets que forman parte de la agrupación y realiza sucesivas llamadas a la API de twitter para obtener los retweets de cada uno

c) Error en la base de datos

Acción de los actores	Respuesta del sistema
	4. Se produce un error de lectura/escritura en la base de datos
	5. El sistema muestra una alerta al usuario con el error obtenido

d) Actualización desde la vista de la base de datos

Acción de los actores	Respuesta del sistema
1. El usuario se encuentra en la vista de la visualización de la base de datos	
2. El usuario activa la opción de actualizar retweets de una de las filas de la tabla correspondiente a un tweet/agrupación de tweets	
	3. El sistema procesa la petición de la forma mencionada en el curso normal

Caso de uso	Comenzar tracking
<i>Identificador:</i>	CU - 014
<i>Actores</i>	Usuario
<i>Descripción</i>	El usuario activa la opción de comenzar tracking (o desactivar tracking) para que el sistema comience a actualizar de forma automática un tweet o un grupo de Tweets y actualice periódicamente la base de datos
<i>Prioridad</i>	Primario, esencial
<i>Referencias cruzadas</i>	CU – 009, el tracking se activa desde el panel de análisis

Curso de eventos normal

Acción de los actores	Respuesta del sistema
1. El usuario, desde el panel de análisis activa la opción de comenzar tracking	
	2. El sistema, marca el Tweet analizado en el panel de análisis y lo introduce en la lista de tracking
	3. El sistema periódicamente actualiza los retweets referentes a cada tweet registrado en la lista de tracking

Curso de eventos alternativo

a) Desactivar tracking

Acción de los actores	Respuesta del sistema
1. El usuario, desde el panel de análisis desactiva la opción de comenzar tracking	
	2. El sistema, extrae el Tweet en el listado de tracking de Tweets, dejando de actualizar periódicamente los retweets

Caso de uso	Alertar usuarios
<i>Identificador:</i>	CU - 015
<i>Actores</i>	Usuario, Twitter
<i>Descripción</i>	El usuario, desde el panel de análisis escribe un mensaje de respuesta para que sea enviado a todos los tweets que forman parte del grupo de análisis, alertando de la posible difusión de una noticia falsa.
<i>Prioridad</i>	Secundaria, deseado
<i>Referencias cruzadas</i>	CU – 009, la opción se activa desde el panel de análisis

Curso de eventos normal

Acción de los actores	Respuesta del sistema
1. El usuario desde el panel de análisis edita el mensaje con el que desea alertar al usuario que ha escrito el tweet	
2. El usuario activa la opción de alertar usuario	
	3. El realiza una llamada a la API de Twitter con la petición del usuario
4. El actor Twitter procesa y realiza la petición, y devuelve el resultado de la operación	

	5. El sistema recibe el resultado exitoso de la petición y la muestra al usuario
--	--

Curso de eventos alternativo

a) Error de la API de Twitter

Acción de los actores	Respuesta del sistema
4. El actor Twitter, procesa la petición y se produce un error.	
	5. El sistema recibe el error, y alerta al usuario con el error correspondiente

Caso de uso	Analizar cuenta bot
<i>Identificador:</i>	CU - 016
<i>Actores</i>	Usuario, Botometer
<i>Descripción</i>	El usuario introduce una cuenta en el buscador de bots, y el sistema a través de la API Botometer le indica la probabilidad de que esa cuenta sea un bot.
<i>Prioridad</i>	Secundario, deseado
<i>Referencias cruzadas</i>	CU – 009, la opción se activa desde el panel de análisis

Curso de eventos normal

Acción de los actores	Respuesta del sistema
1. El usuario accede al panel de detección de bots a través del menú	
2. El usuario introduce el nombre de la cuenta que desea analizar	
	3. El sistema realiza una llamada a la API Botometer, enviando la cuenta introducida por el usuario
4. El actor Botometer*, analiza la cuenta de usuario y devuelve los resultados del análisis	

- | | |
|--|---|
| | 5. El sistema procesa los resultados obtenidos y los muestra al usuario a través de la interfaz |
|--|---|

*El actor Botometer forma parte de un sistema externo, que recibe solicitudes a través de una API, y proporciona resultados relativos a la cuenta analizada, entre los que se encuentra un resultado que indica la probabilidad de que la cuenta analizada sea un bot.

El sistema trata este actor como un sistema externo, por lo que no se entra en detalle de los posibles subprocesos que éste utilice, y se muestran a efectos de este sistema como “una caja negra”, a la que se hacen peticiones y se obtienen resultados. Botometer, como se indica en la sección de estado del arte es un sistema ampliamente conocido y utilizado en la comunidad investigadora.

Curso de eventos alternativo

a) Error de conexión vía API con Botometer

Acción de los actores	Respuesta del sistema
2. El usuario introduce el nombre de la cuenta que desea analizar	
	3. El sistema realiza una llamada a la API Botometer, enviando la cuenta introducida por el usuario
	4. Se produce un error en el proceso de comunicación con Botometer
	5. El sistema informa al usuario del error

5.4. Matriz de trazabilidad de casos de uso

La matriz de trazabilidad de casos de uso relaciona los casos de uso definidos con los diferentes requisitos del sistema. Esta matriz permite conocer la consistencia del modelo y para validar que los casos de uso creados contemplan los requisitos definidos.

Además, si se sigue una metodología de desarrollo basada en casos de uso permite conocer los requisitos que engloba cada uno de los casos, ayudando en la priorización del desarrollo y en el establecimiento de las diferentes iteraciones en metodologías ágiles.

A continuación, se presenta la matriz de trazabilidad, se ha dividido en secciones por cuestiones visuales.

	Registro de nuevo usuario	Recuperación de contraseña	Login de usuario	Editar perfil de usuario	Búsqueda de nuevos tweets	Almacenamiento de tweets	Búsqueda en la base de datos
Requisitos de login y registro							
Registro de usuarios	X						
Recuperación de contraseña		X					
Login de usuarios			X				
Fin de sesión			X				
Visualización de perfil de usuario				X			
Actualización de perfil de usuario				X			
Eliminación de perfil de usuario				X			
Requisitos de obtención de Tweets							
Búsqueda de Tweets por tema				X			
Busqueda de Tweets por usuario				X			
Visualización de Tweets de cuentas relevantes				X			
Requisitos de base de datos - Tweets							
Base de datos - Almacenamiento de nuevos Tweets					X		
Base de datos - Búsqueda de Tweets en la base de datos						X	
Base de datos - Visualización de Tweets almacenados						X	
Base de datos - Eliminación de Tweets almacenados						X	
Requisitos de base de datos - Agrupaciones							
Base de datos - Búsqueda de agrupaciones de Tweets						X	
Base de datos - Visualización de agrupaciones de Tweets						X	
Base de datos - Edición de agrupaciones de Tweets						X	
Base de datos - Eliminación de agrupaciones de Tweets						X	

Ilustración 11. Matriz de trazabilidad (Detalle 1)

	Visualización de cuentas relevantes	Visualización del panel de análisis	Analizar tweet	Actualizar retweets	Búsqueda de tweets similares	Agrupar tweets
Requisitos de base de datos - Cuentas relevantes						
Base de datos - Añadir nueva cuenta relevante	X					
Base de datos - Búsqueda de cuentas relevantes	X					
Base de datos - Visualización de cuentas relevantes	X					
Base de datos - Eliminación de cuentas relevantes	X					
Requisitos de análisis						
Análisis - Vista de panel de análisis		X				
Análisis - Gráfica de difusión de Tweets			X			
Análisis - Tabla cuentas de mayor impacto			X			
Análisis - Tabla primeras cuentas de difusión			X			
Análisis - Mostrar retweets				X		
Análisis - Actualización de retweets				X		
Análisis - Obtención de tweets similares					X	
Requisitos de agrupación de Tweets						
Agrupación - Agrupación de tweets						X
Agrupación - Añadir Tweet a la agrupación						X
Agrupación - Eliminación de tweet de agrupación						X
Agrupación - Visualización de gráfica agregada						X

Ilustración 12. Matriz de trazabilidad (Detalle 2)

	Agrupar tweets	Comenzar tracking	Alertar a usuarios	Analizar cuenta Bot
Requisitos de agrupación de Tweets				
Agrupación - Agrupación de tweets	X			
Agrupación - Añadir Tweet a la agrupación	X			
Agrupación - Eliminación de tweet de agrupación	X			
Agrupación - Visualización de gráfica agregada	X			
Requisitos de tracking				
Tracking - Activar trackeo automático		X		
Tracking - Desactivar trackeo automático		X		
Tracking - Periodicidad de actualización		X		
Tracking - Actualización de status		X		
Requisitos de respuesta				
Respuesta - Crear una respuesta			X	
Respuesta - Validar respuesta			X	
Respuesta - Enviar respuesta			X	
Requisitos de detección de Bots				
Detección de Bots				X

Ilustración 13. Matriz de trazabilidad (Detalle 3)

5.5. Plan de gestión de la calidad

En caso de reproducir el presente proyecto, se recomienda la construcción de un plan de gestión de la calidad que incluya un plan de pruebas y de test de código. No obstante, la forma de asegurar la calidad del proyecto corresponde a cada equipo de acuerdo con sus propias metodologías de trabajo y estándares, así como la tecnología disponible y presupuesto.

6. DISEÑO DEL SISTEMA

6.1. Arquitectura del sistema.

El modelo de implementación del sistema está basado en el famoso modelo de arquitectura de software MVC (Modelo - Vista - Controlador). En la arquitectura MVC, existen tres elementos fundamentales: el modelo, que es el componente que se encarga de la gestión de la información a través de una representación de la misma, y que pone a disposición del controlador cuando este lo requiere; la vista, que es el elemento encargado de presentar la información a través de las interfaces que permiten interactuar al usuario; y el Controlador, que se encarga de la gestión de eventos y de gestionar adecuadamente la relación entre la Vista y el Controlador .

La elección de este sistema se debe a dos razones fundamentales: La primera se debe al framework elegido web2py, un framework completo que está basado en el modelo MVC. Tratar de adaptar la arquitectura a otro tipo de arquitectura supone, no sólo desaprovechar las ventajas que web2py ofrece de por sí, sino añadir complejidad innecesaria al sistema que se desea construir. Por otra parte, la elección de este modelo se basa la facilidad del mismo para su comprensión y para seguir un desarrollo estructurado, donde cada uno de los elementos de la arquitectura tiene una función claramente definida y acotada.

En el framework Web2py, el MVC funciona de la siguiente manera:

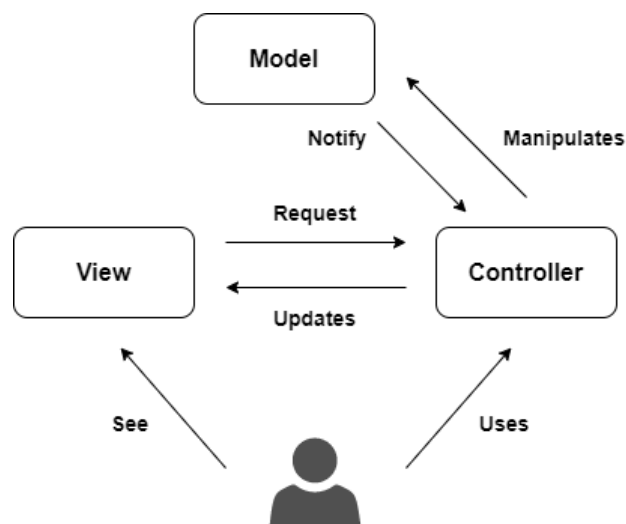


Ilustración 14. Framework Web2py

Como muestra el diagrama, el controlador es el encargado de manipular la estructura de datos del modelo, y a su vez de actualizar las vistas con las que el usuario interactúa. La estructura de archivos que viene predefinida en Web2py contiene (entre otros) los siguientes elementos:

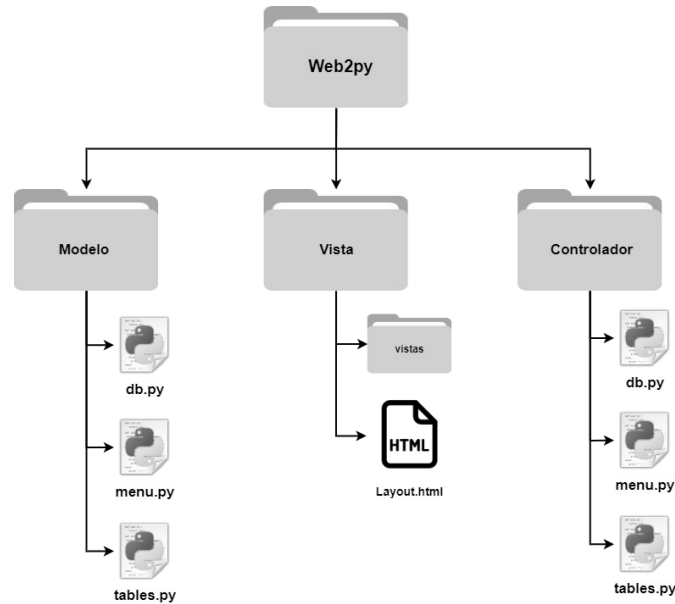


Ilustración 15. Estructura Web2py

Modelo

En el modelo encontramos principalmente tres archivos:

- Menu.py: Las plantillas de las vistas predefinidas en web2py usan un menú en la parte superior que el usuario puede controlar. Este menú se controla con el archivo menu.py
- Tablas: Este archivo contiene la definición de las tablas de la base de datos. Cada nueva tabla o modificación de una tabla existente de la base de datos se hace a través de este script.
- Db.py: Encargado de manejar la capa de abstracción que permite la interacción con la base de datos.

Vista

En el directorio de vistas podemos encontrar:

- Layout.html: Ejemplo de plantilla sobre el que se pueden implementar el resto de las vistas.
- Default: Directorio que contiene cada una de las vistas a las que el cliente tiene acceso.

Controlador

- api.py: Este elemento actúa de como una API, recogiendo las llamadas del cliente y procesándolas en el servidor. Cuando se recibe una petición desde el cliente se hace a esta api, donde un método gestiona todo el proceso y se comunica con el modelo si es necesario para el almacenamiento de datos en la base de datos.
- default.py: Este archivo contiene la definición de las tablas de la base de datos. Cada nueva tabla o modificación de una tabla existente de la base de datos se hace a través de este script.

6.2. Entorno operacional

El diagrama que se presenta a continuación expone desde un punto de vista conceptual la forma en la que funciona y se relaciona el sistema con los agentes. Como se observa en el diagrama, en el sistema el controlador abarca una importante carga de tareas, puesto que es a través de la api del controlador desde donde se gestiona el flujo de información entre las vistas, el modelo y cada uno de los sistemas externos. Además, las vistas que se presentan al usuario son llamadas en los archivos default.py que controla el menú de usuario, y api.py que controla la comunicación cliente-servidor.

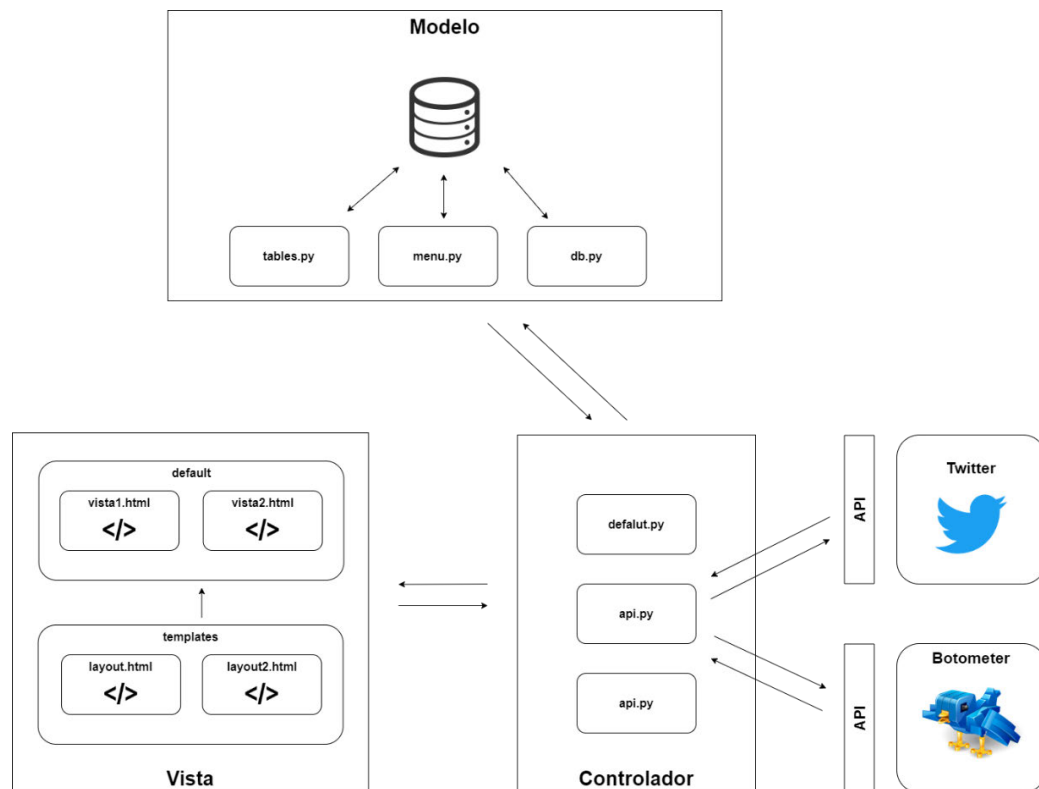


Ilustración 16. Entorno operacional

6.3. Definición del modelo de datos

A continuación, se muestra el modelo de información sobre el que se basa la construcción del sistema, que posteriormente se desglosará en diferentes partes para su análisis.

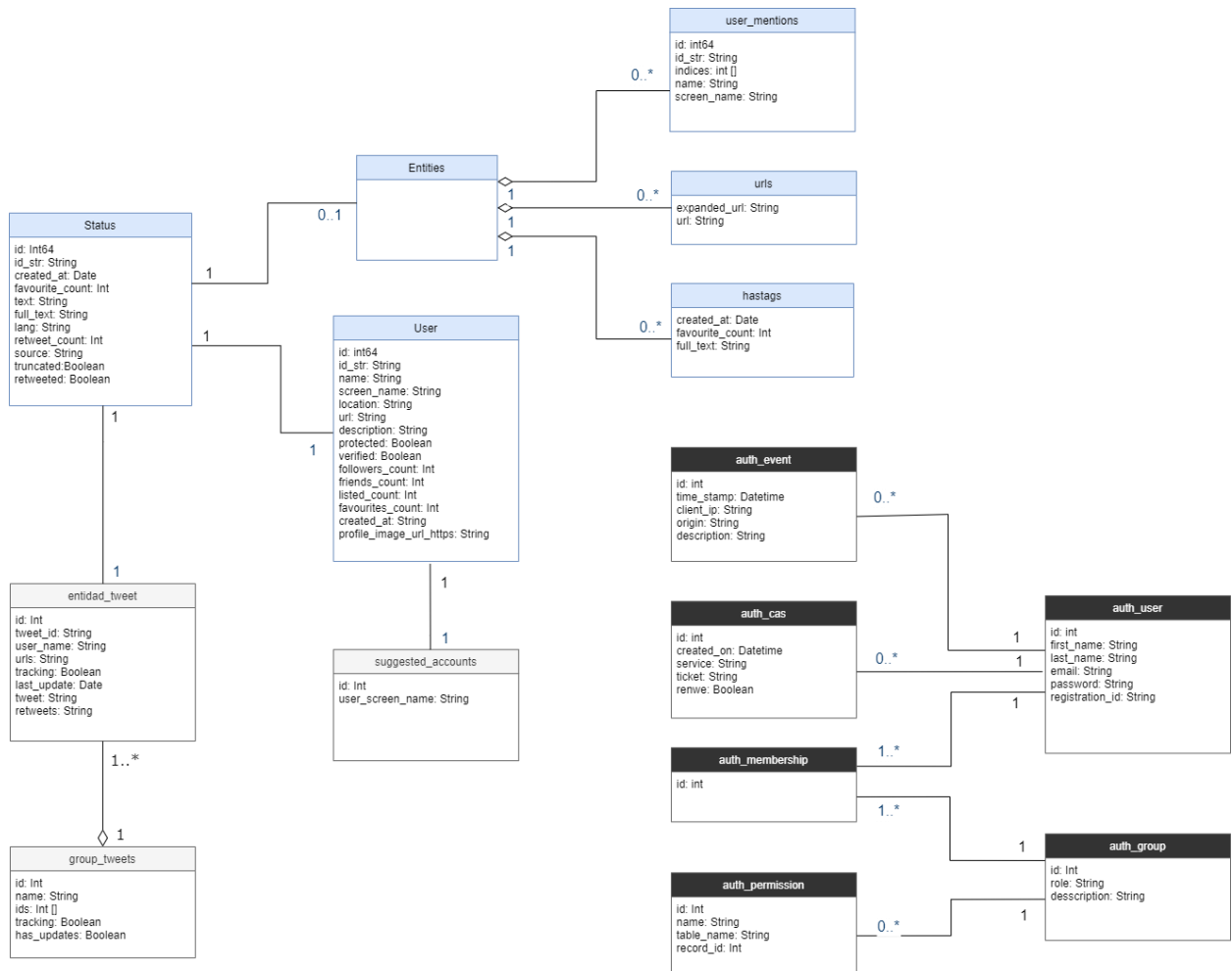


Ilustración 17. Modelo de datos

En el modelo de información se distinguen tres módulos por colores. En azul se encuentran representado el modelo de datos creado para la representación de aquellos datos que provienen de Twitter. En negro el modelo que viene preestablecido en el framework Web2py que se encarga de todo el sistema de autenticación y gestión de usuarios y roles. Si bien no se han hecho uso de algunas de ellas

como de los roles de usuario, ya vienen predefinidas en el sistema. Por último, en gris claro la representación de las clases del sistema que no pertenecen a ninguna de las anteriores categorías.

Modelo de datos de Twitter

Para la definición del modelo de datos, es necesario conocer primero el modelo de datos establecido en Twitter, ya que la aplicación web se basa en la capacidad de conectarse con su API para traer y tweets y otra información relevante.

Cuando se hace una llamada a la API para “Obtener un tweet”, ésta te devuelve un objeto JSON que a su vez contiene otros objetos o entidades. Para tener una idea clara de cómo es ese objeto y los elementos que contiene se presenta el siguiente esquema:

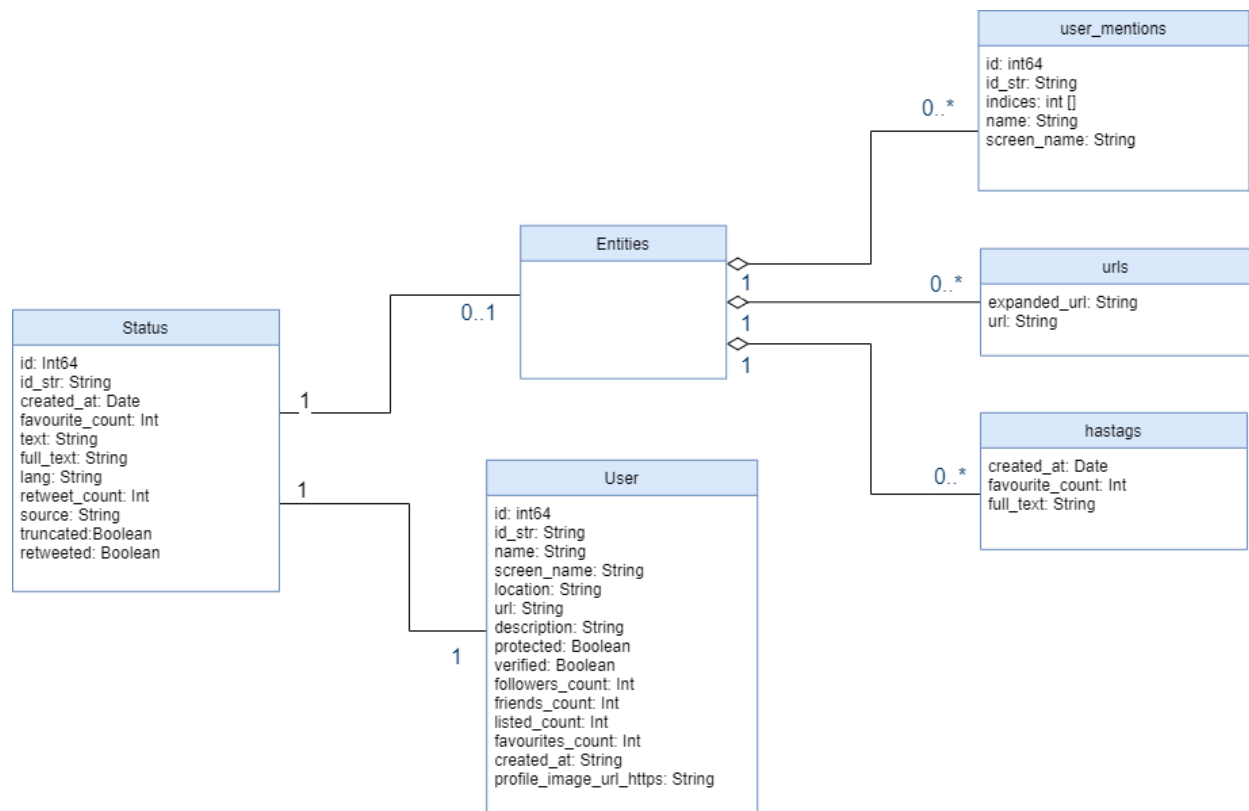


Ilustración 18. Modelo de datos Twitter

Es importante mencionar que en este esquema no se encuentran todos los elementos que componen el modelo de datos de Twitter, y de los que se exponen no se han incluido todos los atributos y entidades del modelo ya que muchos son opcionales y no son relevantes para esta versión del sistema. Para más detalle sobre cada atributo y objeto que compone el modelo se puede visitar la documentación oficial.

Como elementos principales para este proyecto (de los mencionados anteriormente) podemos observar los siguientes, de los que se detallan sus atributos principales:

Status, también conocido comúnmente como Tweet, es uno de los elementos principales en los que se basa Twitter ya que es el elemento que se comparte

Status	
id: Int64 id_str: String created_at: Date favourite_count: Int text: String full_text: String lang: String retweet_count: Int source: String truncated: Boolean retweeted: Boolean	<p>id: id en formato de número</p> <p>id_str (El objeto status contiene dos id, uno llamada id y otro llamado id_str. Entre estos dos, el id con el que se debe trabajar es id_str. Esto es lo que recomienda la propia documentación de Twitter, puesto que id es un número entero que debido a su longitud puede causar problemas al tratarse en JavaScript.</p> <p>created_at: fecha de creación</p> <p>full_text: texto del tweet.</p> <p>lang: Lenguaje del tweet</p> <p>retweet_count: número de veces que ha sido retuiteado</p> <p>favourite_count: número de veces que ha sido marcado como favorito</p> <p>source: la fuente de procedencia del tweet</p>

Tabla 5 Tabla de modelo de datos - Status

Como muestra el diagrama de relaciones, la clase **Status** se relaciona con la clase **User** y con la clase **Entities**, de manera que un **Status** contiene siempre un **User** y puede contener 0 ó 1 **Entities**.

User
id: int64 id_str: String name: String screen_name: String location: String url: String description: String protected: Boolean verified: Boolean followers_count: Int friends_count: Int listed_count: Int favourites_count: Int created_at: String profile_image_url_https: String

id: id en formato de número

id_str: De forma análoga al status Tweet expuesto anteriormente, el usuario cuenta con un identificador.

name: El nombre escogido por el usuario

url: Url al perfil de usuario

description: Descripción que aparece en el perfil de usuario

screen_name: El nombre de la cuenta (@nombre), también es identificador del usuario.

followers_count: El número de seguidores del usuario

favourites_count: Número total de tweets que el usuario de la cuenta ha marcado como favoritos

created_at: Fecha y hora de creación de la cuenta en formato UTC

profile_image_url: La imagen de cuenta de usuario

Tabla 6. Tabla de modelo de datos - Usuario

Entities

Las entidades son objetos que contienen los Status que a su vez contienen elementos contextuales comunes en Twitter, como son (tal y como se definen en twitter) hashtags, urls, media, user_mentions, polls y symbols.

user_mentions
id: int64 id_str: String indices: int [] name: String screen_name: String

hashtags
created_at: Date favourite_count: Int full_text: String

urls
expanded_url: String url: String

Ilustración 19. Entidades de twitter

User_mentions: contiene cada una de las menciones que se encuentran en el cuerpo del Tweet.

Hashtags: con cada uno de los hashtags incluidos en el tweet

Urls: las posibles urls que contiene el tweet, en formato corto (url) y expandido (expanded_url)

Modelo de datos propio

Sobre la representación del sistema de Twitter, trabaja el modelo de datos propio definido principalmente en tres clases

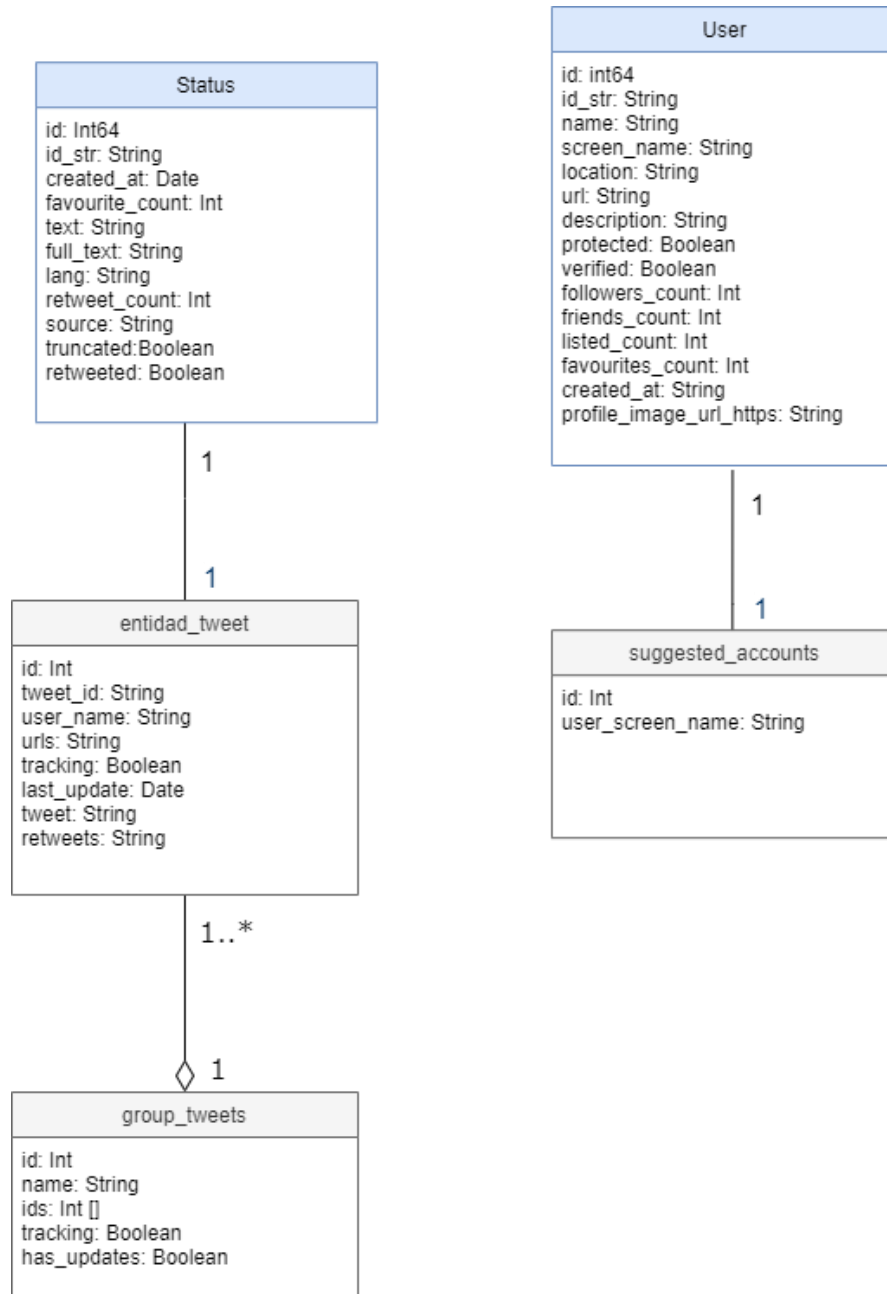
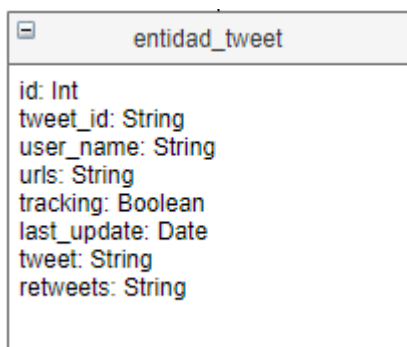


Ilustración 20. Modelo de datos propio

entidad_tweet: Esta clase es probablemente una de las más importantes del sistema, ya que sirve como nexo de unión entre la representación de información de Twitter y la propia del sistema. Contiene los siguientes elementos:



id: número identificador único e incremental
tweet_id: el id del tweet al que hace referencia en formato string
user_name: el nombre de la cuenta que ha creado el Tweet
urls: la url extraída de la propiedad Entities que posee el Tweet en caso de que la tuviera
tracking: un indicador para conocer si un Tweet del sistema está siendo trackeado o no
last_update: la última fecha de actualización
tweet: El objeto que se extrae de Twitter se guarda en forma de texto como un atributo entero de la clase entidad_tweet. * Ver la justificación técnica al respecto que hay a continuación.
retweets: La lista de retweets que se extrae de Twitter con respecto a un tweet se guarda en forma de texto como un atributo entero de la clase entidad_tweet. *Ver la justificación técnica al respecto que hay a continuación.

Tabla 7. Modelo de datos propio - entidad tweet

NOTA: Si bien en un diagrama de clases no se representan como atributos las demás clases con las que una está relacionada, en este diagrama es necesario reflejarlo por la forma en la que se relacionan las clases status y entidad_tweet. A continuación, se expone la justificación técnica

JUSTIFICACIÓN TÉCNICA

Es importante mencionar, que el sistema recupera tweets a través de la API de Twitter, y se los muestra al cliente a través de las vistas. En todo momento el tweet y los elementos que contiene (usuario, retweets, urls, etc) funcionan como un bloque único, por lo que la forma de almacenarlos en la base de datos es convirtiendo ese objeto status de formato JSON a texto, y guardándolo como un bloque entero (En lugar de desagregarlo por cada uno de los atributos)

Será en las vistas cuando usando Vue.js se recupere ese objeto JSON y se trate como una clase, accediendo a las entidades que contiene (usuario, urls, etc). Trabajar de esta forma simplifica mucho el

modelo de datos del sistema, puesto que no es necesario crear en la base de datos tablas que simulen el modelo de datos de Twitter, desagregando cada Tweet en los diferentes elementos que lo contienen.

Esta decisión se ha tomado principalmente por dos motivos: el primero, por rapidez de acceso a la base de datos. Obtener en una única llamada a la base de datos el objeto entero evita tener que reconstruir los objetos a partir de diferentes tablas haciendo varias llamadas a sucesivas tablas. En segundo lugar, por simplicidad. Sabiendo que los objetos se van a tratar como un bloque y enviarse directamente a las vistas, crear más tablas en la base de datos sólo añade complejidad al sistema y ninguna funcionalidad extra. Lo mismo ocurre con el atributo retweets, que contiene la lista con todos los retweets pertenecientes a un tweet.

Suggested_accounts representa las cuentas relevantes seguidas por el usuario

suggested_accounts
user_screen_name: String

User_screen_name: contiene el nombre único de la cuenta de usuario de Twitter

Group_tweets representa una agrupación de entidad_tweet relacionados

group_tweets
id: Int name: String ids: Int [] tracking: Boolean has_updates: Boolean

name: contiene el nombre del grupo

main_id: id del tweet principal sobre el que se creó la agrupación

ids: lista de ids de los tweets que conforman una agrupación

tracking: atributo que refleja si un grupo de tweets está en estado de tracking

has_updates: atributo que refleja si hay nuevos retweets detectados

retweets: La lista de retweets que se extrae de Twitter con respecto a un tweet se guarda en forma de texto como un atributo entero de la clase entidad_tweet.

*Ver la justificación técnica al respecto que hay a continuación.

Tabla 8. Modelo de datos propio - agrupación de tweets

Access control

El propio framework web2py facilita el desarrollo en muchos aspectos, uno de los principales es el hecho de que el entorno proporciona un módulo de control de acceso predefinido que proporciona la capacidad de crear roles de usuarios, registros a la aplicación y controles de administrador. Esto se consigue gracias a la clase Auth. A continuación, se expone el modelo de datos del control de acceso. Dado que es un módulo predefinido del framework, no es necesario ningún tipo de implementación y ya viene preparado para su uso. El módulo de control de acceso está representado de la siguiente manera:

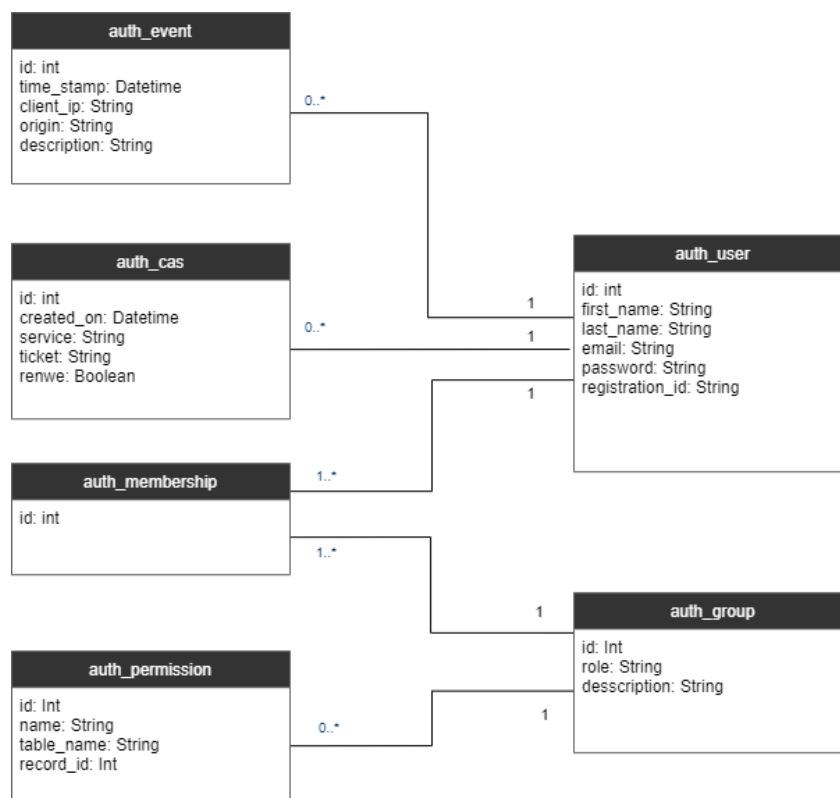


Ilustración 21. Modelo de datos Web2py

En este proyecto no se está haciendo uso de los roles de usuario ni de la aplicación de permisos a usuarios según su rol, por lo que actualmente son clases en desuso, pero se mantienen en el modelo de datos para reflejar las posibilidades de escalabilidad del sistema.

7. DISEÑO DE LA INTERFAZ

En este apartado se presenta el diseño del sistema a nivel de interfaz. La forma recomendable para poder explicar el diseño de las interfaces del sistema a posibles usuarios o a los propios desarrolladores del sistema es a través de prototipos.

Mediante el uso de estos prototipos se pueden exponer de forma sencilla las principales funcionalidades del sistema, y obtener feedback en una etapa temprana del proyecto tratando de definir el sistema más adecuado antes de comenzar a desarrollar (evitando posibles costes futuros tras una implementación errónea).

Para ello en este proyecto, previamente a desarrollar la solución final se ha creado un prototipo a bajo nivel, donde se pone el foco en las funcionalidades y las características más que en el propio diseño, colores, estilos, etc. Tras obtener feedback y tras varias iteraciones con un prototipo de bajo nivel (si fuera necesario), se podría proceder a realizar un prototipo de alto nivel, donde se hace foco en el diseño, estilos, colores con el fin de conseguir una buena experiencia de usuario y una interfaz gráfica de mayor calidad.

Es importante mencionar que a pesar de que el prototipo trata de reflejar el producto final, es posible (incluso probable) que el software resultante sufra cambios con respecto al prototipo, debido a nuevas implementaciones o a restricciones del sistema.

Nota: Con esta sección del proyecto, no se pretende limitar a futuros desarrolladores o proyectos similares que deseen reproducir el sistema, sino guiar el desarrollo del software mediante prototipos ya validados. De hecho, en caso de querer reproducirse este mismo sistema se invita a los desarrolladores a crear nuevos prototipos y diseños diferentes que mejoren o amplíen los que aquí se presentan.

7.1. Prototipo de bajo nivel

Durante el proyecto se han realizado varios prototipos a bajo nivel a partir de una pizarra, que más adelante se han digitalizado utilizando software creado específicamente para ello. En concreto se ha usado la herramienta Balsamiq.

Mediante la herramienta, se han creado las principales pantallas de la aplicación web, que se reflejan en un gráfico a tamaño reducido a continuación. Puede observarse el detalle ampliado en el anexo

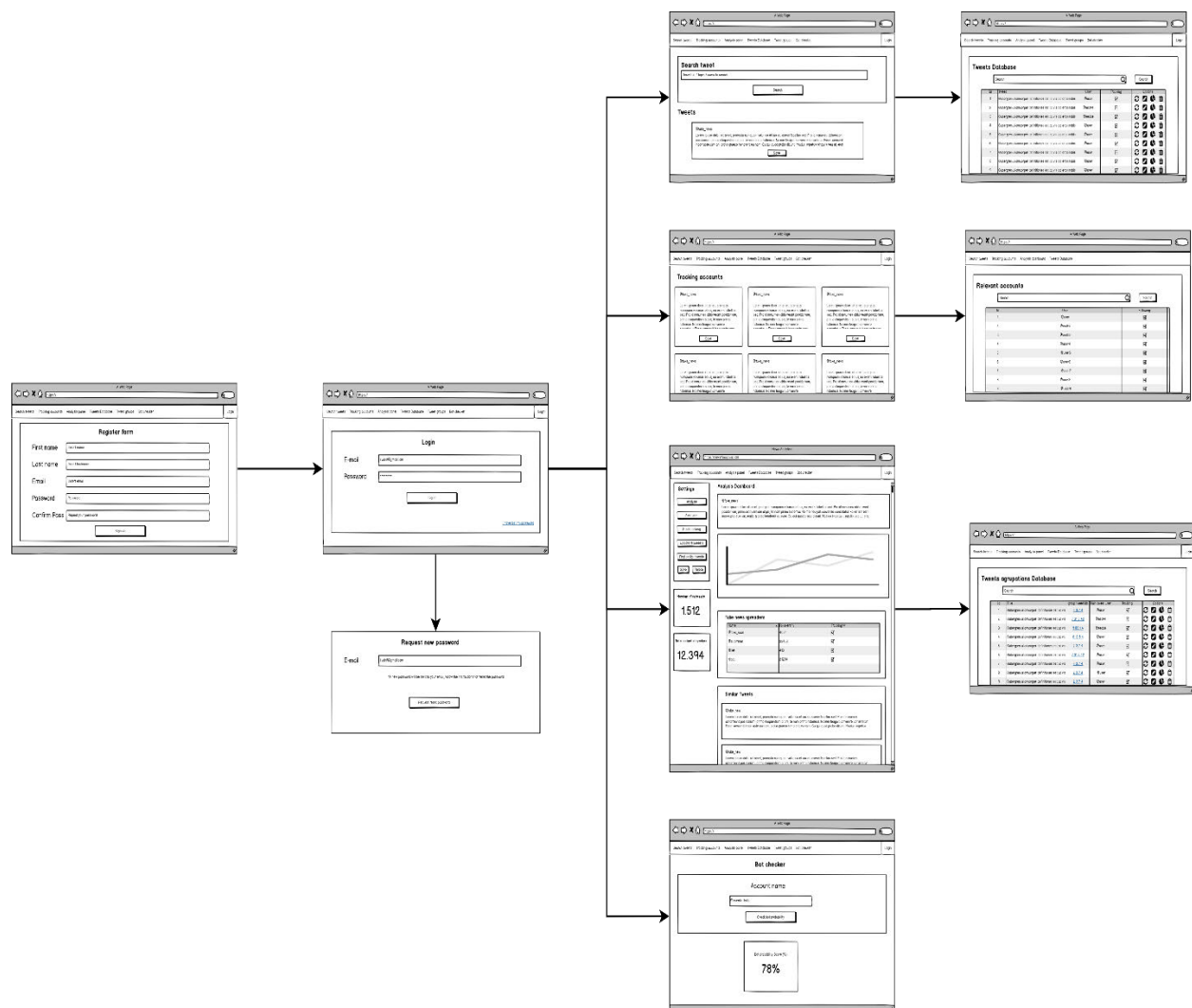
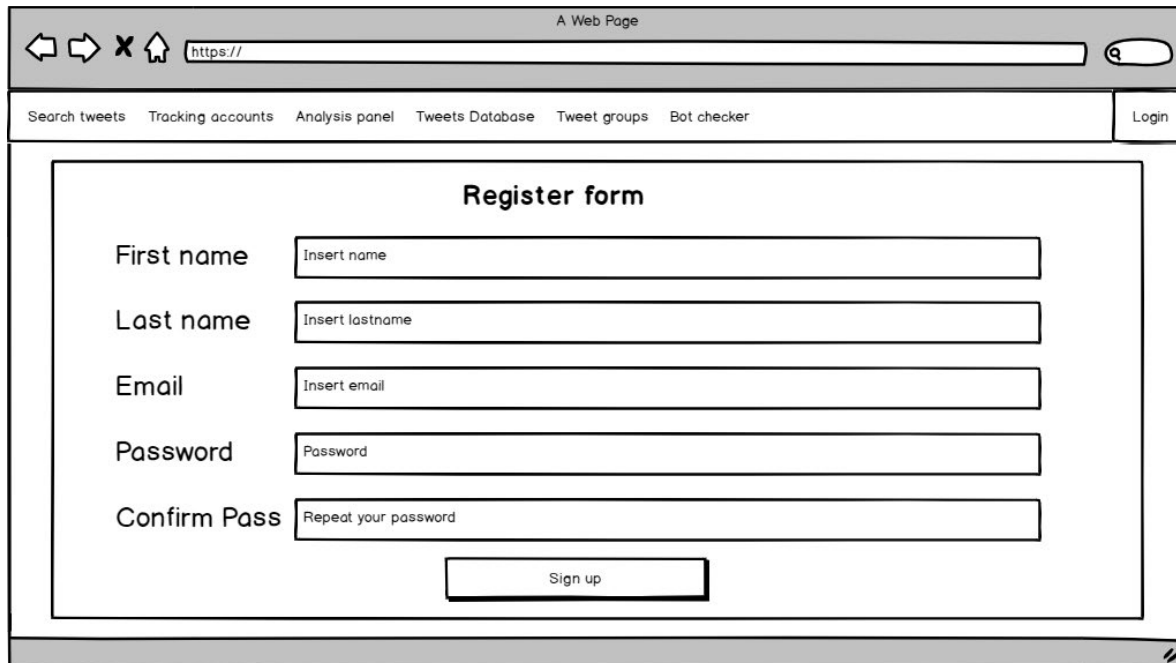


Ilustración 22. Flow de la aplicación

A continuación, se muestra el detalle ampliado de cada una de las interfaces que compone el sistema:

Registro de usuarios

La primera pantalla del sistema es la que permite el registro de nuevos usuarios, mediante los campos de nombre, apellido, email y contraseña



A Web Page

Search tweets Tracking accounts Analysis panel Tweets Database Tweet groups Bot checker Login

Register form

First name

Last name

Email

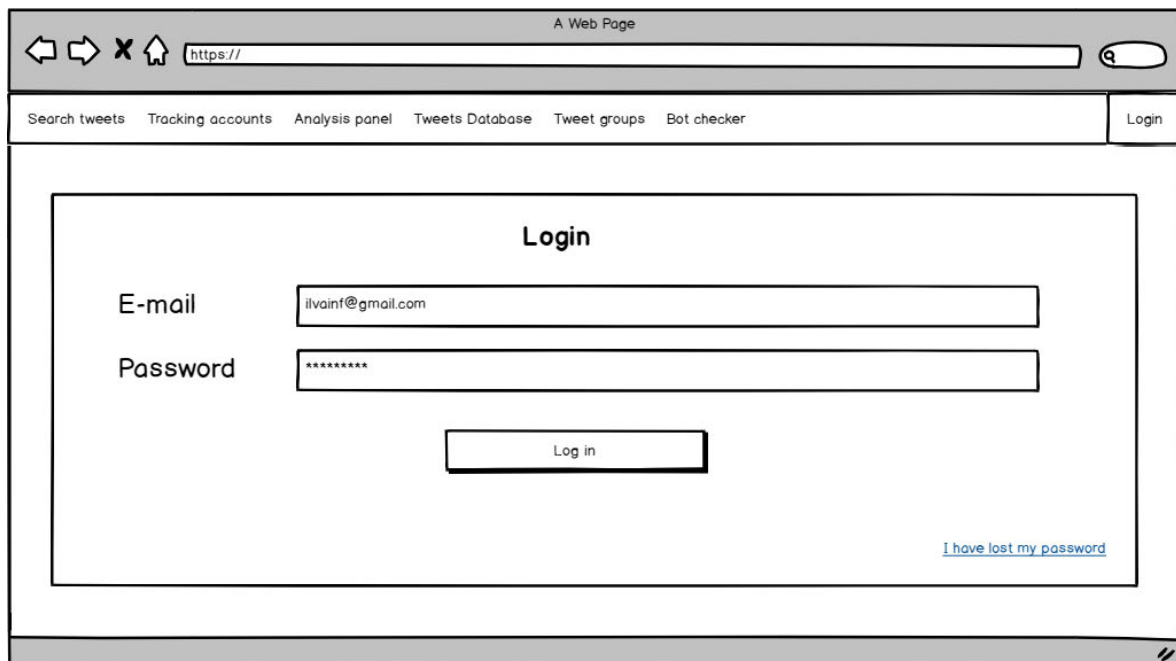
Password

Confirm Pass

Ilustración 23. Prototipo a bajo nivel, Pantalla de registro

Pantalla de login

Cualquier usuario registrado, puede acceder al sistema mediante la pantalla de login de usuarios, introduciendo su correo y contraseña



A Web Page

Search tweets Tracking accounts Analysis panel Tweets Database Tweet groups Bot checker Login

Login

E-mail

Password

[I have lost my password](#)

Ilustración 24. Prototipo a bajo nivel, Pantalla de Login

Si el usuario lo necesita, tiene a su disposición un link de “He perdido mi contraseña”, tras el que se abre un pop-up para solicitar una nueva vía email

Request new password

E-mail

*A new password will be sent to your email, follow the instructions for reset the password

Ilustración 25.. Prototipo a bajo nivel, recuperación de contraseña

Barra de menú

A través del menú de opciones, se puede acceder a cada una de las funcionalidades reflejadas en las pantallas que se muestran a continuación. Además cuenta con un acceso rápido para el login/Registro de usuarios

Search tweets Tracking accounts Analysis panel Tweets Database Tweet groups Bot checker

Ilustración 26. Prototipo a bajo nivel, menú principal

Buscador de Tweets

Esta pantalla permite, a través de un buscador en la aplicación, buscar sobre un tema, un link, el título de una noticia o un usuario, y el sistema muestra los tweets resultado de la búsqueda a través de la interfaz. Estos tweets pueden ser guardados mediante un botón.

The wireframe shows a web browser window titled 'A Web Page'. The address bar contains 'https://'. Below the browser window is a navigation bar with links: 'Search tweets', 'Tracking accounts', 'Analysis panel', 'Tweets Database', 'Tweet groups', 'Bot checker', and a 'Login' button on the right. The main content area is divided into two sections. The first section, titled 'Search tweet', contains a text input field with the placeholder 'Insert url / topic / users to search' and a 'Search' button below it. The second section, titled 'Tweets', contains a tweet card for '@fake_news'. The tweet text is a block of Lorem Ipsum: 'Lorem ipsum dolor sit amet, prompta numquam volumus et ius, eu exerci fabellas sed. Pro id nonumes abhorreant posidonium, prima eloquentiam at pri, te nam prima habemus. No mei feugait convenire consetetur. Facer senserit incorrupte cum an, oratio graeco hendrerit eu nam. Cu qui quod probo dicunt. Mucius impetus singulis sea ut, erat'. Below the tweet text is a 'Save' button. The bottom of the browser window shows a double-slash icon.

Ilustración 27. Prototipo a bajo nivel, Pantalla búsqueda de tweets

Panel de análisis

En esta pantalla se reúnen la mayoría de las características que se mencionan en el apartado de alcance del sistema y engloban el panel de análisis de tweets y las acciones que se pueden tomar a la hora de analizar el alcance de un tweet. En el panel se muestran gráficas, tablas y datos relevantes sobre la difusión del tweet.

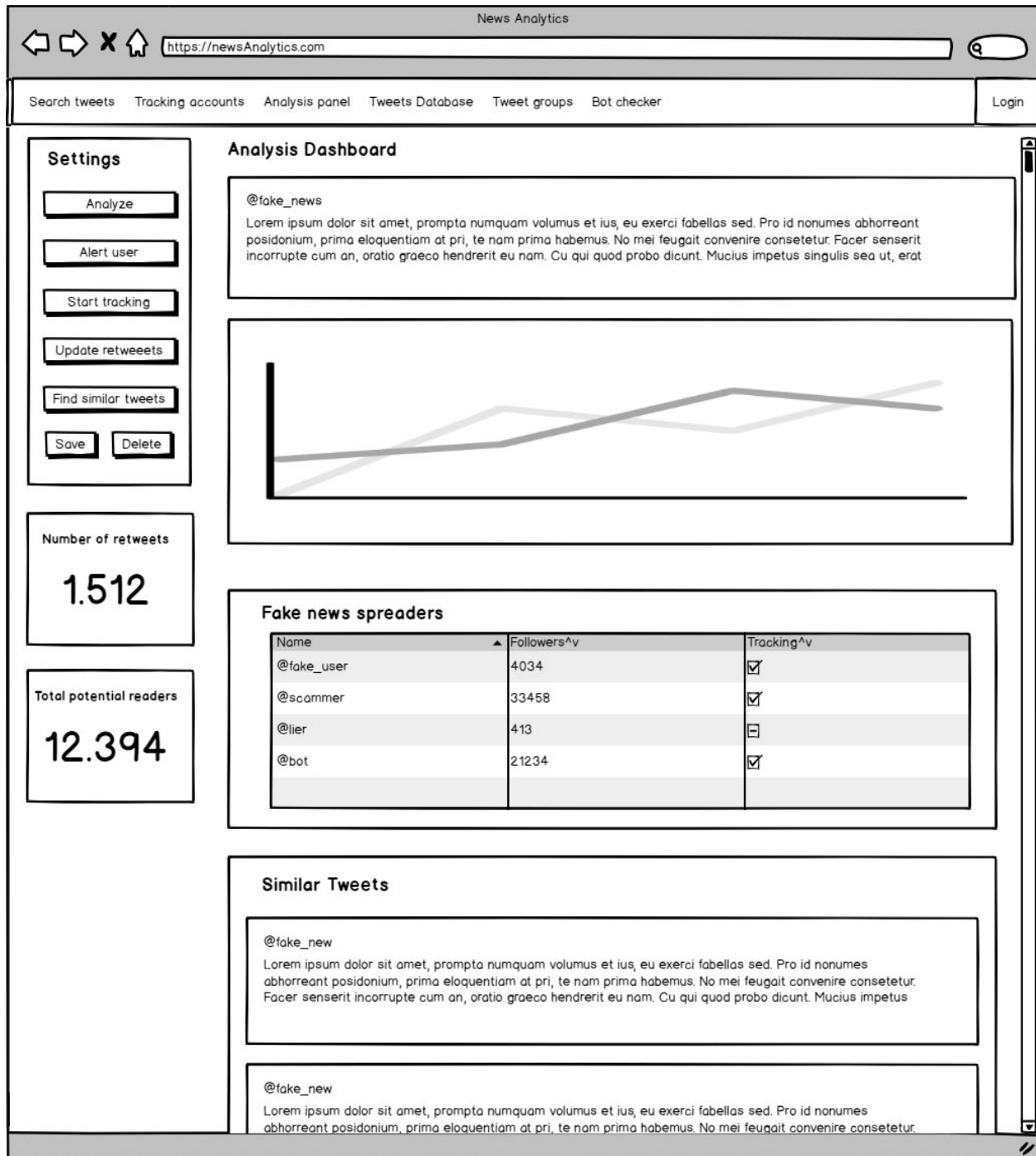


Ilustración 28.. Prototipo a bajo nivel, Panel de análisis

Pantalla de cuentas relevantes

En esta pantalla aparecen los últimos tweets posteados por las cuentas que el usuario considera relevantes. De los Tweets mostrados, el usuario puede guardar cualquiera de ellos en la base de datos

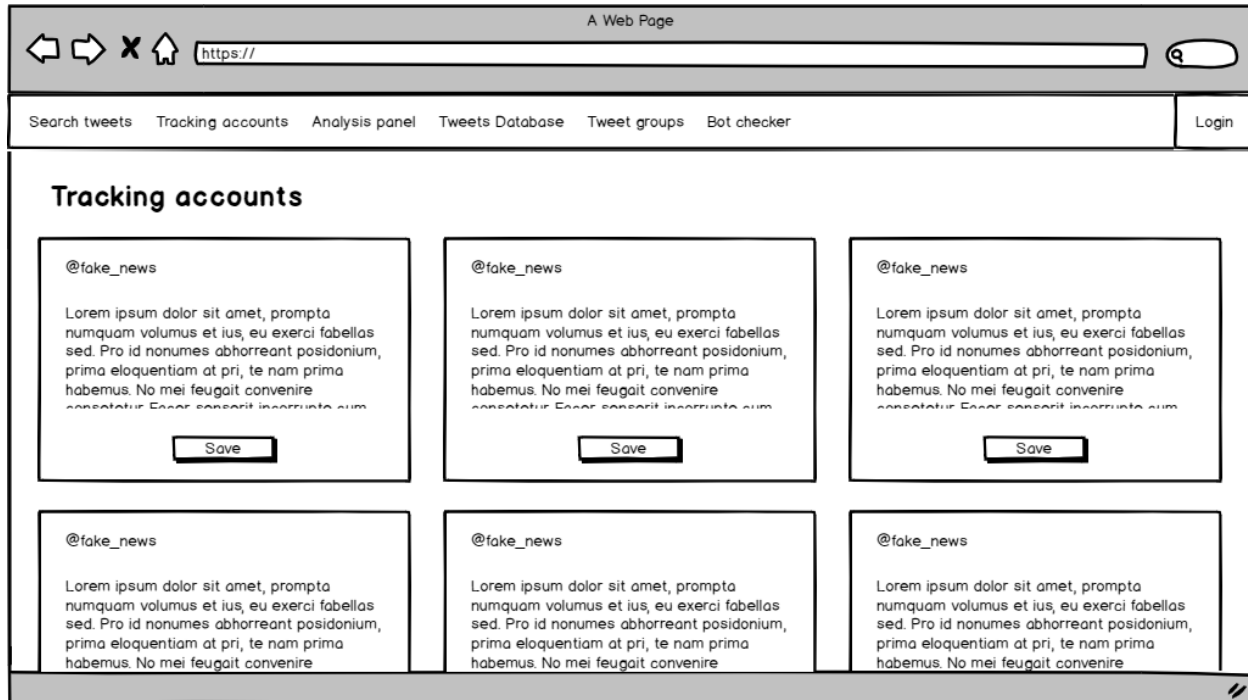


Ilustración 29. Prototipo a bajo nivel, Pantalla de cuentas relevantes

Base de datos, buscador de cuentas relevantes

En esta pantalla se puede hacer una búsqueda en la base de datos de aquellas cuentas relevantes que el usuario está siguiendo, y marcar la opción de “unfollow” para dejar de seguirlas.

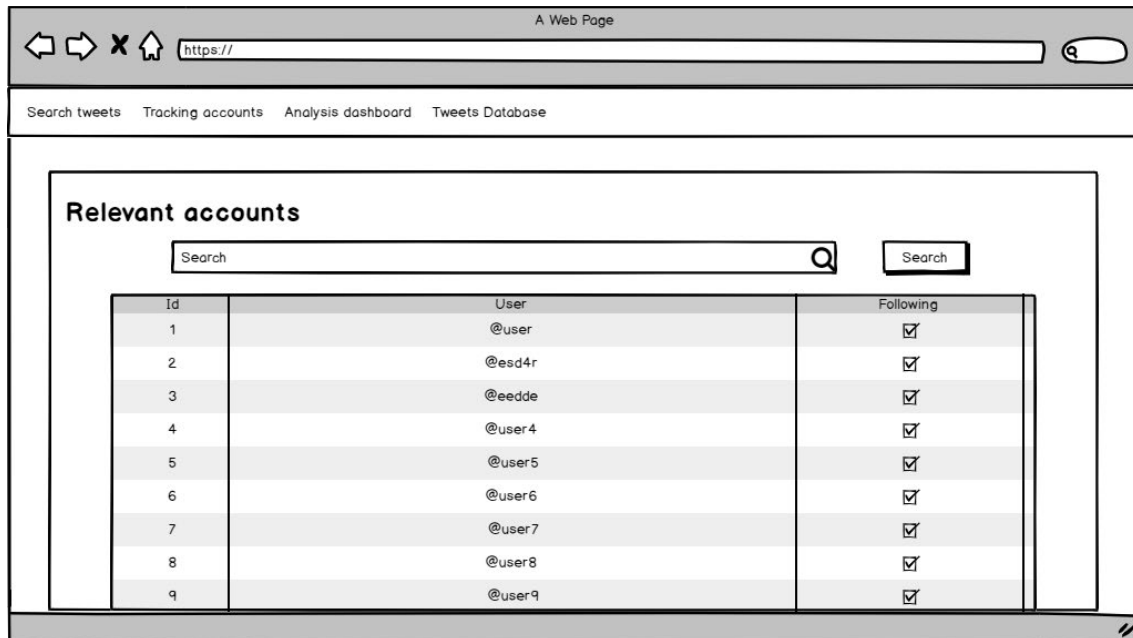


Ilustración 30.. Prototipo a bajo nivel, Base de datos - cuentas relevantes

Base de datos, buscador de tweets

En esta pantalla es posible visualizar la base de datos donde están los tweets almacenados, y acceder a cualquiera de ellos a través de una búsqueda. Además de cada tweet almacenado se ofrecen las opciones de ver, actualizar, editar, analizar y eliminar.



Ilustración 31.. Prototipo a bajo nivel, Base de datos – búsqueda de tweets

Buscador de agrupaciones de tweets

En esta pantalla el usuario puede hacer una búsqueda en la base de datos de las diferentes agrupaciones de Tweets que tiene creadas, así como las opciones para activar/desactivar el tracking, actualizar, analizar la agrupación, editarla o eliminarla

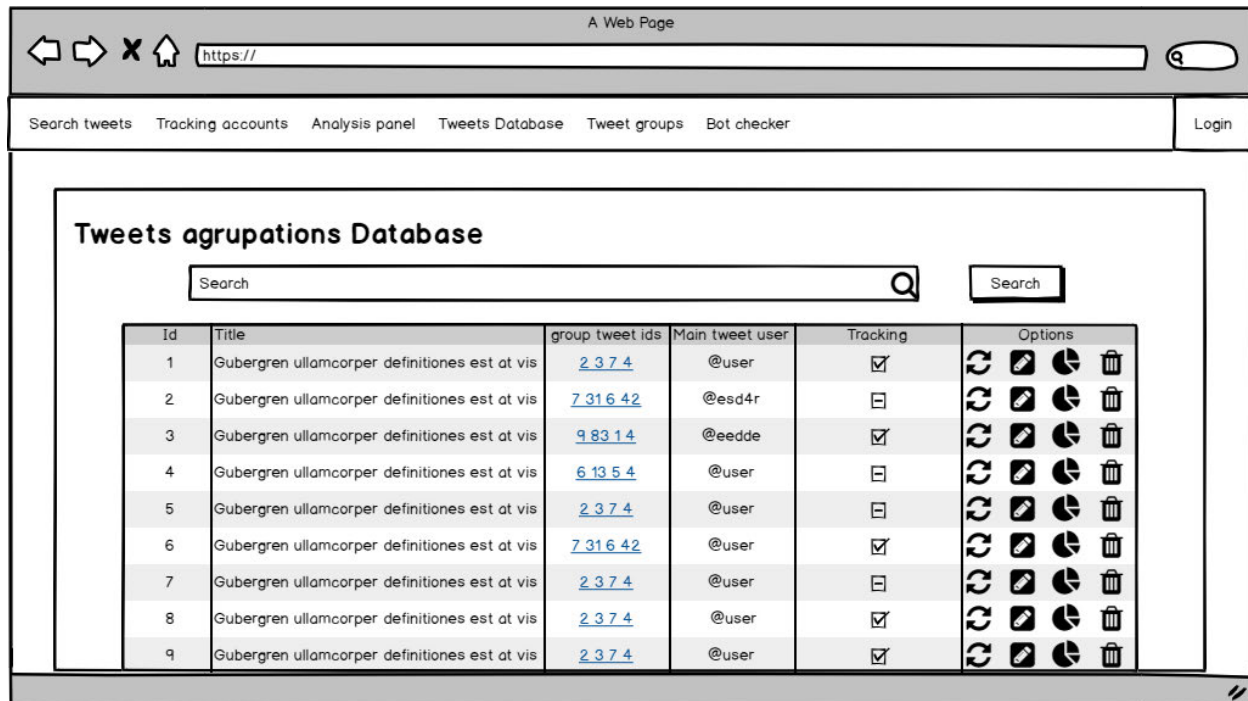


Ilustración 32.. Prototipo a bajo nivel, Base de datos - búsqueda de agrupaciones de tweets

Resultados del prototipo a bajo nivel

Gracias al prototipado de bajo nivel se ha recogido el siguiente feedback de diferentes fuentes y se ha incluido nuevas características en el sistema. A continuación, se expone el feedback obtenido:

- Tweets: Tratar de hacer que los tweets se parezcan más a la interfaz de Twitter, para que sean fácilmente reconocibles para el usuario. Mostrar otros elementos del tweet como la foto del usuario, el nombre, la fecha y el número de retweets.
- Buscador: Incluir la posibilidad de un buscador avanzado que permita buscar por hashtag, usuario, fechas... etc.
- Bots: Indicar si un usuario es un bot o no
- Generales: Incluir más iconos de diferentes tipos para que hagan la aplicación más atractiva

7.2. Panel de diseño o Mood Board.

Para el desarrollo final del sistema, se ha diseñado un panel que recoge los principales estilos, iconos y colores con los que se desea trabajar. Este panel servirá como guía a la hora de construir las interfaces, tratando de seguir un desarrollo coherente entre las diferentes interfaces, tratando de mejorar la experiencia del usuario cuando usa el sistema. Debido al objetivo del software relacionado con las fake news y con Twitter, se desea:

Por un lado, transmitir la seriedad y la importancia del tema que se trata, utilizándose estilos serios que alejen la atención del software en si, y lo centren en la utilidad del mismo. Por ello se evitarán el uso de colores llamativos y se centrarán los estilos en colores oscuros y elegantes.

Por otro lado, al ser los propios tweets que contienen las noticias el foco de atención, se desea seguir un estilo similar a Twitter, que permita al usuario relacionar de forma inconsciente los elementos que encuentra en la interfaz con el propio Twitter, utilizando la misma gama de colores y estilos similares a la hora de mostrar un tweet en la interfaz, pero tratando de evitar ser una copia exacta e incluyendo diseños propios que permitan una ligera diferenciación.

Por último, se ha incluido el icono propio del sistema, que identificará la herramienta y se podrá visualizar en diferentes partes como en el título de la página o en la barra de marcadores del navegador.

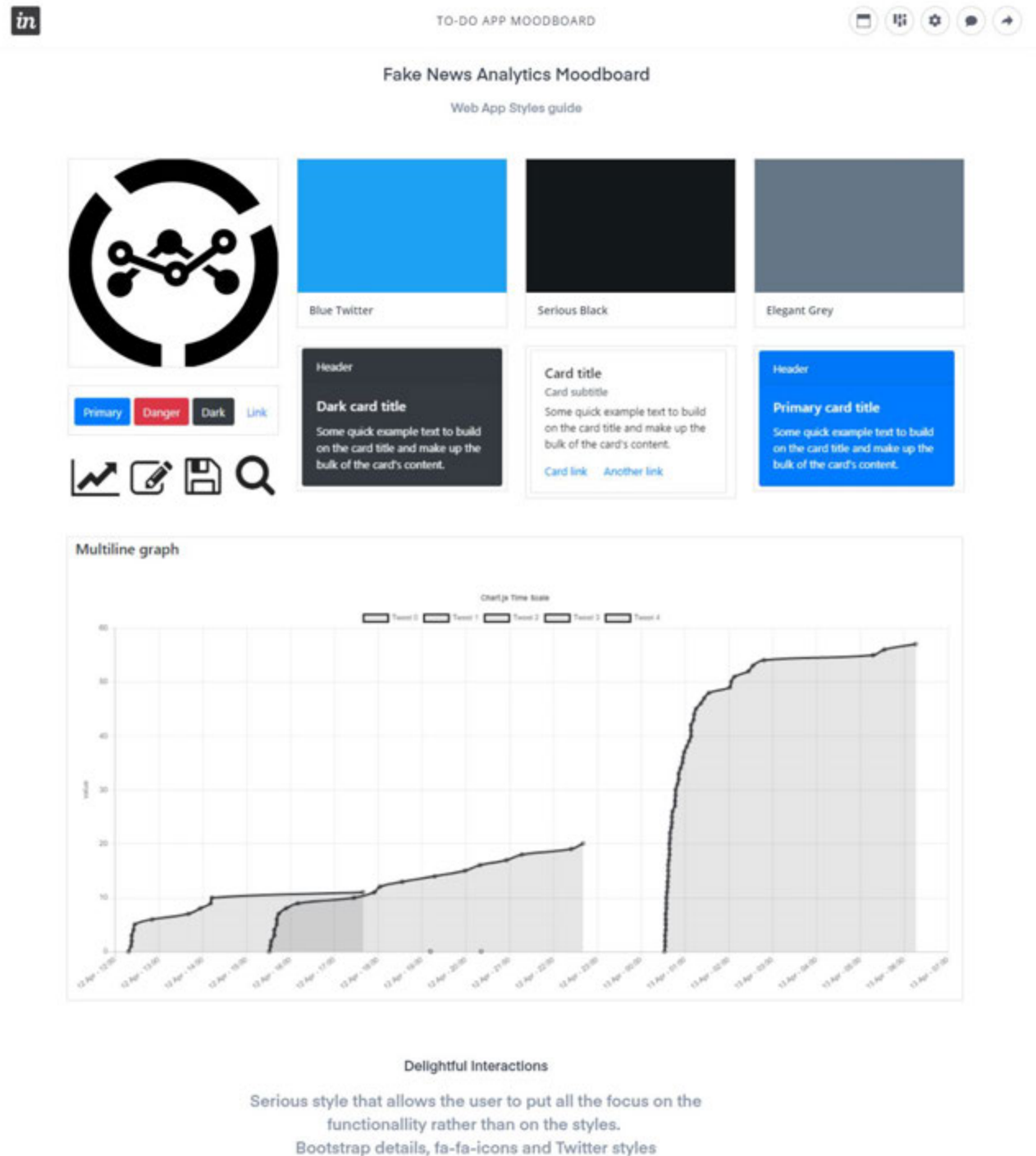


Ilustración 33. Diseño del sistema, Moodboard

7.3. Tecnologías utilizadas

A continuación, se detallan las principales tecnologías necesarias para el desarrollo del proyecto. Como se ha mencionado anteriormente, la aplicación está basada en un modelo cliente-servidor, por lo que cabe diferenciar las tecnologías usadas en el lado del cliente (front-end) de las tecnologías usadas en el lado del servidor (back-end)

7.3.1. Tecnologías front end

HTML, CSS y Javascript

Para el diseño de las vistas del cliente, se usa la famosa combinación de las tres tecnologías **HTML**, **CSS** y **JavaScript**. Mediante esta combinación se consigue crear una interfaz con unos diseños y estilos serios pero atractivos a la vista para el usuario, con unas interfaces sencillas de usar.

Además de las mencionadas, es importante destacar el uso de dos tecnologías que han aportado gran valor en el lado front-end, de cara a mejorar las funcionalidades de la aplicación y a los estilos de la misma.

Vue.js

Vue.js tal y como se define en su documentación oficial es un “*framework Javascript progresivo para construir interfaces de usuario*”. Debido a que el sistema gira en torno a unas entidades que son análogas a los tweets, se debe buscar una forma de desarrollo que permita reproducir el mismo componente en repetidas ocasiones en la interfaz de usuario.

Una de las principales características de Vue es que permite la creación de componentes reutilizables a lo largo de la misma o de varias vistas, además de funcionalidades muy atractivas como usar bucles o condicionales para modificar la forma en la que el código HTML presenta los elementos. Esto nos permite crear un elemento como puede ser un tweet, y reproducir el mismo elemento varias veces en la vista cambiando el contenido.

Justificación de la elección

Si bien para construir el sistema no es necesario el uso de esta tecnología, si es recomendable el uso de un framework como el que se presenta u otro similar como React o Angular para conseguir resultados similares. Tras leer algunos estudios comparativos entre los mencionados frameworks, se encontró que Vue presenta un crecimiento exponencial y se ha situado como el framework más popular

de acuerdo a los datos de popularidad en GitHub. A continuación, se presenta una gráfica que refleja el crecimiento de esta tecnología durante los últimos años:

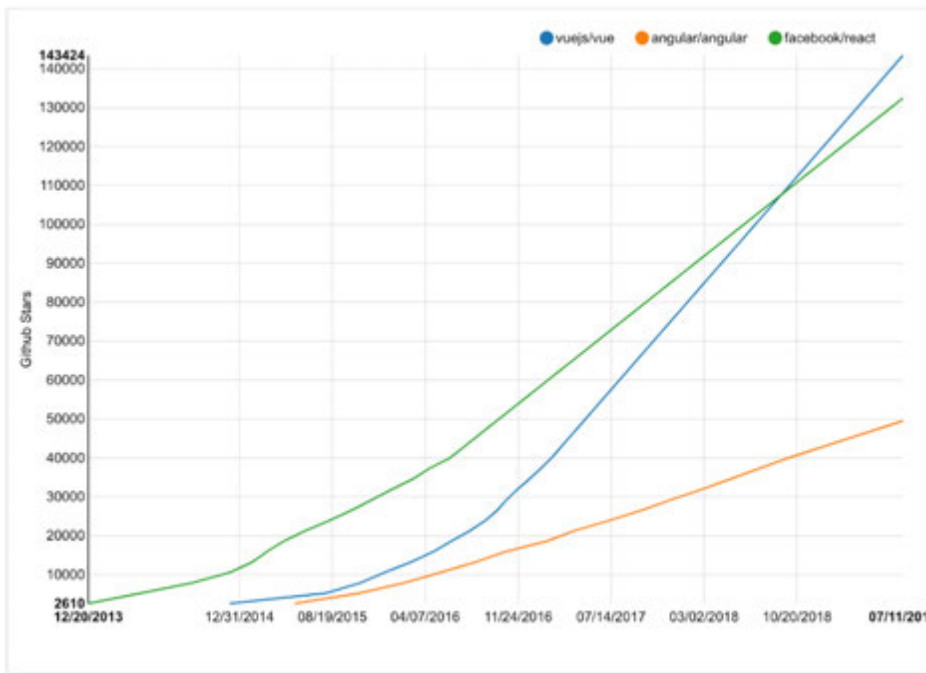


Ilustración 34. Tecnologías utilizadas, justificación de elección Vue.js [33]

7.3.2. Tecnologías back-end

Web2py

Web2py es probablemente la tecnología más destacada que presenta el sistema. La documentación define este sistema como un framework gratuito open source para el desarrollo rápido de aplicaciones web data-driven en Python.

La decisión de usar este framework se debe principalmente a los siguientes motivos:

- Rapidez: web2py permite un desarrollo expreso de aplicaciones web, ya que cuenta con algunos módulos predefinidos ya construidos que ofrecen funcionalidades básicas sobre las que partir. Como ejemplo principal web2py cuenta con sistema de autenticación de usuarios construido o con una base de datos accesible mediante una capa de abstracción
- Sencillez: el hecho de que el back-end esté basado en Python, supone una gran facilidad a la hora de tratar con datos de Twitter, principalmente porque existen librerías predefinidas que permiten la conexión y extracción de datos de Twitter de una manera simplificada.

- Seguridad: la Open Web Application Security Project es una organización sin ánimo de lucro centrada en el estudio y la mejora de la seguridad del software. Esta organización produce un listado de fallos de seguridad comunes, entre los que principalmente se encuentran Ataque por inyección SQL, pérdida de autenticación o ataque XSS (Cross Site Scripting Attack). Web2py cuenta con diferentes medidas de seguridad para estos y otros tipos de vulnerabilidades comunes. A modo de ejemplo: la capa de abstracción de la base de datos impide la manipulación directa y controla cualquier tipo de ataque por inyección. El sistema de autenticación de web2py controla y gestiona la autenticación de usuarios, además de usar SHA-512 para la encriptación de datos y contraseñas. Se puede obtener más información sobre estas y otras medidas de seguridad en su documentación oficial [34].

No obstante, si se desea reproducir este sistema, se recomienda valorar también el uso de otras tecnologías, ya que web2py es un sistema que, si bien sigue funcionando, ha comenzado a quedarse obsoleto.

7.3.3. Otras tecnologías

Git & GitHub

Como principal tecnología para llevar un control de versiones adecuado y facilitar el desarrollo y el progreso del código y de las diferentes versiones, se ha elegido la combinación de git + github. Como principal ventaja git es el uso de branches o ramas, que permiten crear vías de trabajo paralelas evitando la posibilidad de romper el estado actual del código, y llevando un desarrollo estructurado. Github por su parte ofrece un repositorio gratuito para alojar el código de forma remota.

Python-twitter

Python-twitter es una librería Python que facilita la conexión con la API de Twitter. Esta librería viene con una serie de métodos predefinidos que facilitan algunas de las funcionalidades del sistema, como la búsqueda de tuits, búsqueda de cuentas o las contestaciones a mensajes. Puede encontrarse más información en su documentación oficial

8. CONSTRUCCIÓN / RESULTADOS

A lo largo de esta sección se presentan como resultado obtenido del proyecto la solución que se ha construido, que reúne la mayoría de las características expuestas en el diseño. Además, se exponen las limitaciones del proyecto y las posibles vías de explotación económica del proyecto.

8.1. Resultados

A lo largo de este apartado, se exponen el resultado final de la aplicación construida. La versión construida del sistema reúne las principales características del mismo que permiten validar una primera versión de la idea, aunque hay ciertas características en relación con la automatización que no se han llegado a completar. A nivel visual se han seguido las propuestas mostradas en el apartado de diseño, aunque sigue habiendo aspectos que son mejorables.

No obstante, el objetivo principal era contar con un mínimo producto viable que permitiese validar la idea, lo cual sí se ha conseguido. A continuación, se muestran las diferentes secciones de la aplicación.

Registro de usuarios

Esta pantalla permite el registro de nuevos usuarios en la plataforma, está compuesto de un formulario sencillo que permite al usuario introducir sus datos personales y registrarse. El formulario cuenta con un proceso de validación.

Search term	Search user	Relevant accounts	Stored tweets	Grouped tweets	Check bots	Fake News Panel	LOGIN ▾
-------------	-------------	-------------------	---------------	----------------	------------	-----------------	---------

User

First name

Last name

E-mail

Password

Too short

Confirm Password

Password fields don't match

[Sign Up](#)

Copyright © 2020 Powered by [web2py](#)

Ilustración 35. Resultados, registro de usuarios

User

First name

Last name

E-mail

Password

Confirm Password

[Sign Up](#)

Ilustración 36. Resultados, registro de usuarios (Mobile)

Pantalla de login de usuario

Mediante el formulario de login el usuario puede acceder a la plataforma con su email y contraseña

Search term Search user Relevant accounts Stored tweets Grouped tweets Check bots Fake News Panel LOGIN

User

E-mail

Password

☐ Remember me (for 30 days)

[Log In](#)

Copyright © 2020 Powered by web2py

Ilustración 37. Resultados, pantalla de login

User

E-mail

Password

☐ Remember me (for 30 days)

[Log In](#)

Ilustración 38. Resultados, pantalla de login (Mobile)

Panel de búsqueda

El panel de búsqueda permite recuperar Tweets de dos formas posibles: Mediante palabras clave/link y mediante cuentas de usuario:

The screenshot shows a web interface for searching tweets. At the top is a navigation bar with links: Search term, Search user, Relevant accounts, Stored tweets, Grouped tweets, Check bots, Fake News Panel, and a LOGIN button. Below this is the 'Search panel' with a search bar containing 'fake new', an 'Analyze' button, and a link to 'Advanced Search'. The 'Tweets Source' section displays two tweets. The first tweet is from Donald J. Trump (@realDonaldTrump), mentioning Jim Banks and a link to a Breitbart News article. The second tweet is from Eric Feigl-Ding (@DrEricDing), mentioning a fake driver's license permit. Both tweets include engagement metrics (Retweets and Favs) and a 'Guardar' (Save) button.

Search panel
fake new
Analyze
Advanced Search

Tweets Source

	Donald J. Trump by @realDonaldTrump Jim Banks: New York Times Damaged National Security to Hurt Trump https://t.co/6jT8duFYf via @BreitbartNews Thank you Jim. They are Fake News! Link to Tweet Retweets: 18977 Favs: 55120 Date: Wed Mar 18 13:48:38 +0000 2009	Guardar
	Eric Feigl-Ding by @DrEricDing Forget fake drivers license. A new fake permit is born. Crazy dangerous. https://t.co/sVDEGSyPYP Link to Tweet Retweets: 94 Favs: 222 Date: Sat Jan 10 11:01:33 +0000 2009	Guardar

Ilustración 39. Resultados, pantalla de búsqueda de tweets

Ejemplo de resultados tras la búsqueda por cuenta de usuario, donde se muestran los 15 últimos tweets escritos por la cuenta.

Search term

Search user

Relevant accounts

Stored tweets

Grouped tweets

Check bots

Fake News Panel


LOGIN

Search panel

realDonaldTrump

Search

User timeline



Donald J. Trump


by @realDonaldTrump

@FoxNews gladly puts up the phony suppression polls as soon as they come out. We are leading in the REAL polls because people are sick & tired of watching the Democrat run cities, in all cases, falling apart. Also, now 98% Approval Rating in the Republican Party. Another 2016!

Retweets: 4514 Favs: 16047

Date: Sun Jul 05 20:47:28 +0000 2020

Guardar



Donald J. Trump

by @realDonaldTrump

RT @realDonaldTrump: On this wonderful day, we celebrate our history, our heroes, our heritage, our flag, and our FREEDOM. Happy Fourth of...

Retweets: 33316 Favs:

Date: Sun Jul 05 17:09:06 +0000 2020

Guardar

Ilustración 40. Resultados, resultados de la búsqueda por usuario

Acceso a la base de datos

La vista de la base de datos permite visualizar en una tabla los Tweets que han sido almacenados, y tomar ciertas acciones con ellos como:

- Ver: Permite ver la entrada en la base de datos
- Editar: Para editar la entrada de la base de datos
- Borrar: Eliminar la entrada de la base de datos

Search term	Search user	Relevant accounts	Stored tweets	Grouped tweets	Check bots	Fake News Panel	Uki
-------------	-------------	-------------------	---------------	----------------	------------	-----------------	-----

Panel Guardados

25 records found

Tweet Id	User	Text	Analysis panel
99	malditobulo	RT @MalditoDato: 🚫 Es falso que un estudio de Oxford dé a España "la puntuación más alta de los países occidentales" en el rigor de la re...	<input type="button" value="View"/> <input type="button" value="Delete"/>
101	okdiario	🔴 #ÚLTIMAHORA El coronavirus no da respiro a España: 16.353 muertos tras sumar 510 en 24 horas https://t.co/BzHbAiqdEZ	<input type="button" value="View"/> <input type="button" value="Delete"/>
102	elmundoes	💎 Casi 160.000 casos confirmados del nuevo #coronavirus en el estado de Nueva York, más que en cualquier país del mundo, y cerca de 8.000 muertos. ¿Por qué este estado de 19 millones de habitantes ha sido tan afectado por la #pandemia? 🤔 https://t.co/wU9YcFtRAq	<input type="button" value="View"/> <input type="button" value="Delete"/>
103	elmundoes	RT @ElMundoOpinion: #Editorial 📰 "Lo que subyace en la escandalosa parálisis del Gobierno no sólo es su negligente falta de previsión, si...	<input type="button" value="View"/> <input type="button" value="Delete"/>
108	malditobulo	@TxemaMarin @luismirrojias 🤔🤔	<input type="button" value="View"/> <input type="button" value="Delete"/>
109	elmundoes	🔴 Isabel Díaz Ayuso anuncia que se van a comenzar a realizar "test selectivos masivos" a los madrileños para conocer la evolución del #coronavirus. https://t.co/FN76S8uOs	<input type="button" value="View"/> <input type="button" value="Delete"/>
110	malditobulo	RT @ElasdeLuis: Este vídeo con bolsas negras de cadáveres no es del Palacio de Hielo de Madrid es Guayaquil (Ecuador) https://t.co/CbPJFRR...	<input type="button" value="View"/> <input type="button" value="Delete"/>
111	ActualidadRT	Gobiernos y grandes empresas abandonan Zoom por sus graves problemas de seguridad: ¿qué tan peligrosa es la exitosa 'app'? https://t.co/NCzAGyq6Tl	<input type="button" value="View"/> <input type="button" value="Delete"/>
112	okdiario	🔴 @Eduardolinda: «Los Pactos de la Moncloa no pueden servir para mantener al culpable de este desaguisado en La Moncloa» https://t.co/0EUpD2D868	<input type="button" value="View"/> <input type="button" value="Delete"/>
113	elmundoes	#Últimaha del #coronavirus 🔴 España registra 619 fallecidos en las últimas 24 horas, el total de fallecidos por el #COVID19 asciende a 16.972. https://t.co/Fdz4wsG3gP	<input type="button" value="View"/> <input type="button" value="Delete"/>
114	elpais_espana	El presidente reconoce en el Congreso errores pero pide unidad y lealtad: "Toda Europa llegó tarde, pero España actuó antes" https://t.co/de7q8h5NqD	<input type="button" value="View"/> <input type="button" value="Delete"/>
131	elpais_espana	Ni en Marbella, ni en Zahara de los Atunes, ni en Tarifa, ni en Chiclana. Tampoco en Valencia, ni en Llanes (Asturias), ni en Galicia, ni en la Costa Brava. Los turistas nunca llegaron https://t.co/9WnPI3reX9	<input type="button" value="View"/> <input type="button" value="Delete"/>
138	malditobulo	RT @RosaKempin: ¿Ha publicado el BOE que te pueden "expropiar" una vivienda? El Gobierno indica que el programa no permite la expropiación...	<input type="button" value="View"/> <input type="button" value="Delete"/>

Ilustración 41. Resultados, base de datos – tweets

Además, el sistema permite una búsqueda avanzada en la base de datos por varios parámetros:

Search term	Search user	Relevant accounts	Stored tweets	Grouped tweets	Check bots	Fake News Panel	Uki
-------------	-------------	-------------------	---------------	----------------	------------	-----------------	-----

Panel Guardados

25 records found

Tweet Id	User	Text	Analysis panel
99	malditobulo	RT @MalditoDato: 🚫 Es falso que un estudio de Oxford dé a España "la puntuación más alta de los países occidentales" en el rigor de la re...	<input type="button" value="View"/> <input type="button" value="Delete"/>

Ilustración 42. Resultados, base de datos - detalle de búsqueda

Panel de análisis

Este es probablemente el elemento más importante de todo el sistema, es el panel que contiene las principales funcionalidades del sistema:

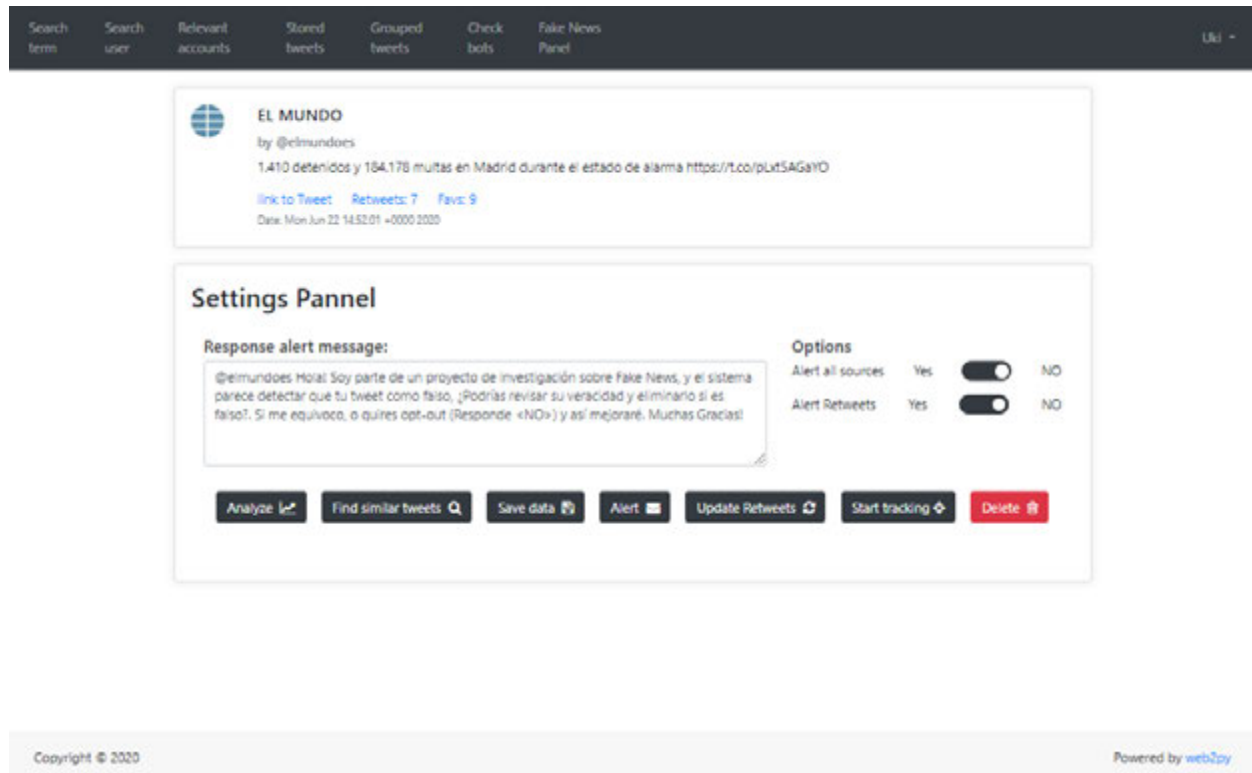


Ilustración 43. Panel de análisis

Cuando desde este panel, se pulsa el botón de analizar, se muestran una serie de gráficas y datos. En la parte principal se muestran los gráficos de difusión en el tiempo (tanto simple como agregado en caso de que haya una agrupación de tweets). A continuación de los gráficos, se muestran las tablas que contienen el top 10 usuarios con mayor alcance potencial (mayor número de seguidores) y el top 10 usuarios por orden de difusión.

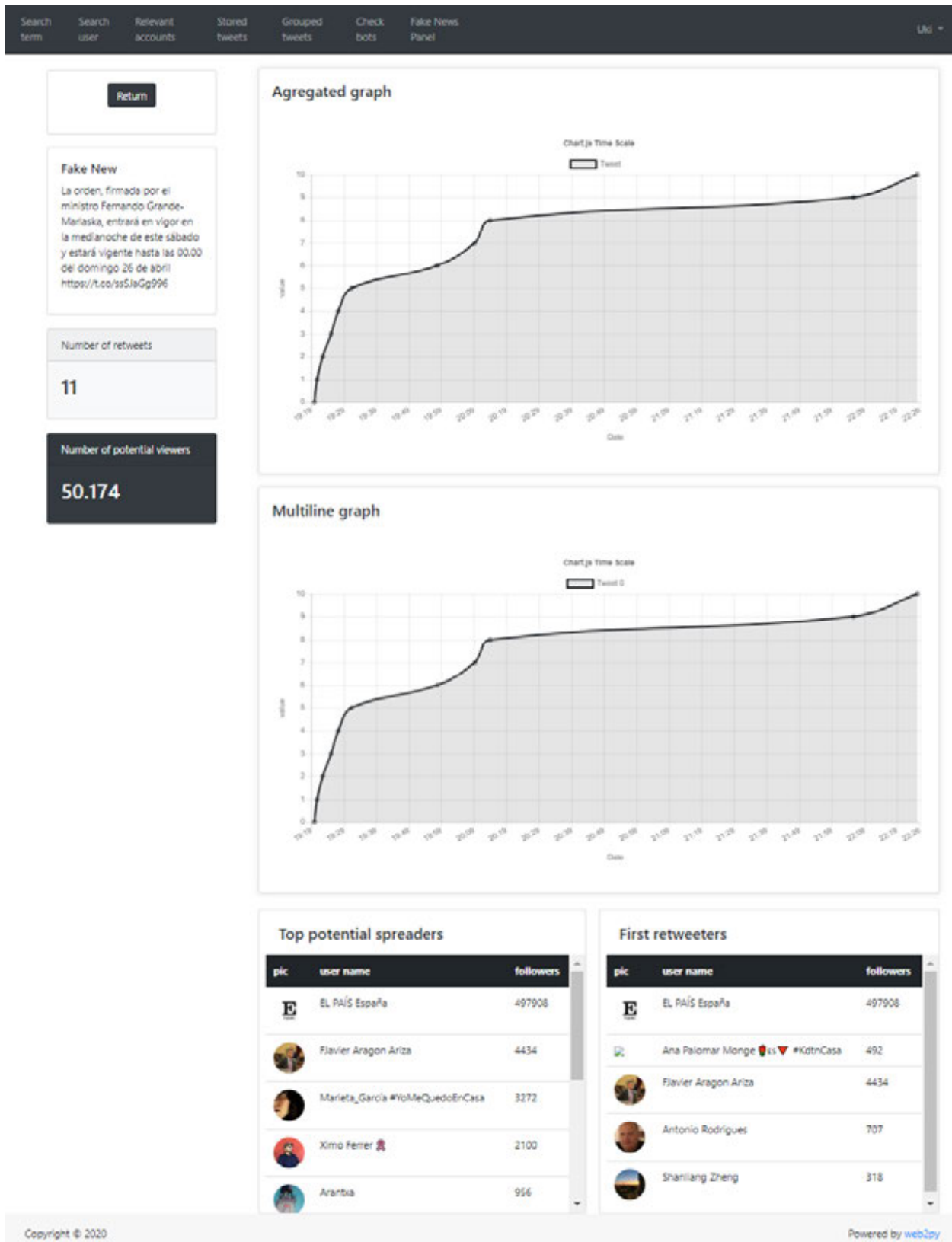



Ilustración 44. Panel de análisis - Resultados del análisis

Ilustración 45. Resultados de análisis

Encontrar Tweets Similares

Al activar la opción de encontrar Tweets similares, la herramienta recoge la url embebida en el Tweet y recupera la página web a la que pertenece. Posteriormente, mediante técnicas de web scrapping extrae el título de la noticia y mediante el uso de la API de Twitter, extrae tweets que contengan el mismo título. Posteriormente los muestra en la herramienta:

Search term
Search user
Relevant accounts
Stored tweets
Grouped tweets
Check bots
Fake News Panel
Log out









EL PAÍS España
by @elpais_espana
La encuesta no prevé la desmovilización por el miedo a la pandemia que tanto temen los partidos, especialmente las formaciones que aspiran a acabar con 11 años de mandatos del PP <https://t.co/X0ixdMupgc>
[Link to Tweet](#) [Retweets: 3](#) [Favs: 5](#)
Date: Sun Jul 05 16:58:23 +0000 2020

Settings Pannel









Response alert message:

@elpais_espana Hola! Soy parte de un proyecto de investigación sobre Fake News, y el sistema parece detectar que tu tweet como falso. ¿Podrías revisar su veracidad y eliminarlo si es falso? Si me equivoco, o quieres opt-out (Responde «NO») y así mejoraré. Muchas Gracias!

Options
Alert all sources Yes ☒ NO
Alert Retweets Yes ☒ NO


Analyze  Find similar tweets  Save data  Alert  Update Retweets  Start tracking  Delete 


Similar tweets

pic	user name	text	followers	favs	Options
	ARM	Fejód lograría su cuarta mayoría absoluta el 12-J https://t.co/Aud3yS1A2R a través de @el_pais	964	31178	
	Joaquín A. de Toledo	Estabilidad en Galicia. Me alegro. Fejód lograría su cuarta mayoría absoluta el 12-J https://t.co/R15C5AkAAs a través de @el_pais	1639	181	
	Ramon Planet es. #liberalismo	Fejód lograría su cuarta mayoría absoluta el 12-J vía @el_pais https://t.co/VL8gPzTbI	372	28546	
	Pablo Arnaldos	Nuestro estudio sobre las elecciones gallegas del próximo domingo para @el_pais https://t.co/08ix2N48oN	144	4	

Other Tweet Sources

This section displays similar tweets founded (other initial sources of the same fake news)


Iñigo Cainzos Achirica
by @icainzos93
Fejód lograría su cuarta mayoría absoluta el 12-J <https://t.co/2unireCph1> vía @el_pais
[Link to Tweet](#) [Retweets:](#) [Favs:](#)
Date: Sun Jul 05 12:00:32 +0000 2020


Mariano Cejas
by @marianocejassr
Fejód lograría su cuarta mayoría absoluta el 12-J | España | EL PAÍS <https://t.co/w6UFRnZaZ>
[Link to Tweet](#) [Retweets:](#) [Favs:](#)
Date: Sun Jul 05 09:28:07 +0000 2020




Puerto de la Cruz
by @puertocruz05
2020-07-05 El País Tag Elecciones Encuesta Encuestas. El PP de Alberto Núñez Fejód se encamina hacia su cuarta mayoría


Ilustración 46. Encontrar tweets similares

Una vez recuperados los Tweets, se pueden agrupar mediante las opciones que tiene cada uno de ellos a la derecha (Agregar, eliminar o guardar como un Tweet independiente). Al ir agregando varios Tweets se añaden a una tabla dinámica que muestra la “agrupación de Tweets”:

Settings Pannel

Response alert message:

@elmundoes Hola! Soy parte de un proyecto de investigación sobre Fake News, y el sistema parece detectar que tu [tweet](#) como falso, ¿Podrías revisar su veracidad y eliminarlo si es falso? Aquí tienes la fuente <https://t.co/KK6LA8Z1jp>

Options

Alert all sources Yes ☐ NO

Alert Retweets Yes ☐ NO

Analyze

Find similar tweets

Save data

Alert

Update Retweets

Start tracking

Delete

Similar tweets

pic	user name	text	followers	scope	Options
	Mallorca Informa	El Mundo: 1.410 detenidos y 184.178 multas en Madrid durante el estado de alarma https://t.co/yWAHN9gPZE @elmundobaleares	729	1377	
	América Hoy	1.410 detenidos y 184.178 multas en Madrid durante el estado de alarma https://t.co/oKQwZgXALc	584	1410	
	El Mundo España	1.410 detenidos y 184.178 multas en Madrid durante el estado de alarma https://t.co/IN8pPSkP8U	22140	47	

Ilustración 47. Agrupación de tweets

Una vez se dispone de una agrupación de Tweets, se pueden analizar como conjunto de datos, de la misma forma que el análisis de un Tweet individual. En esta ocasión, el gráfico agregado muestra la suma de cada uno de los Tweets de forma conjunta, y el gráfico multilínea muestra cada uno de ellos por separado. En la parte inferior se muestran dos tablas, una con los usuarios potencialmente más influyentes a la hora de distribuir la noticia (ordenados de mayor a menor número de seguidores) y la segunda con los primeros usuarios que han distribuido la noticia en orden temporal:

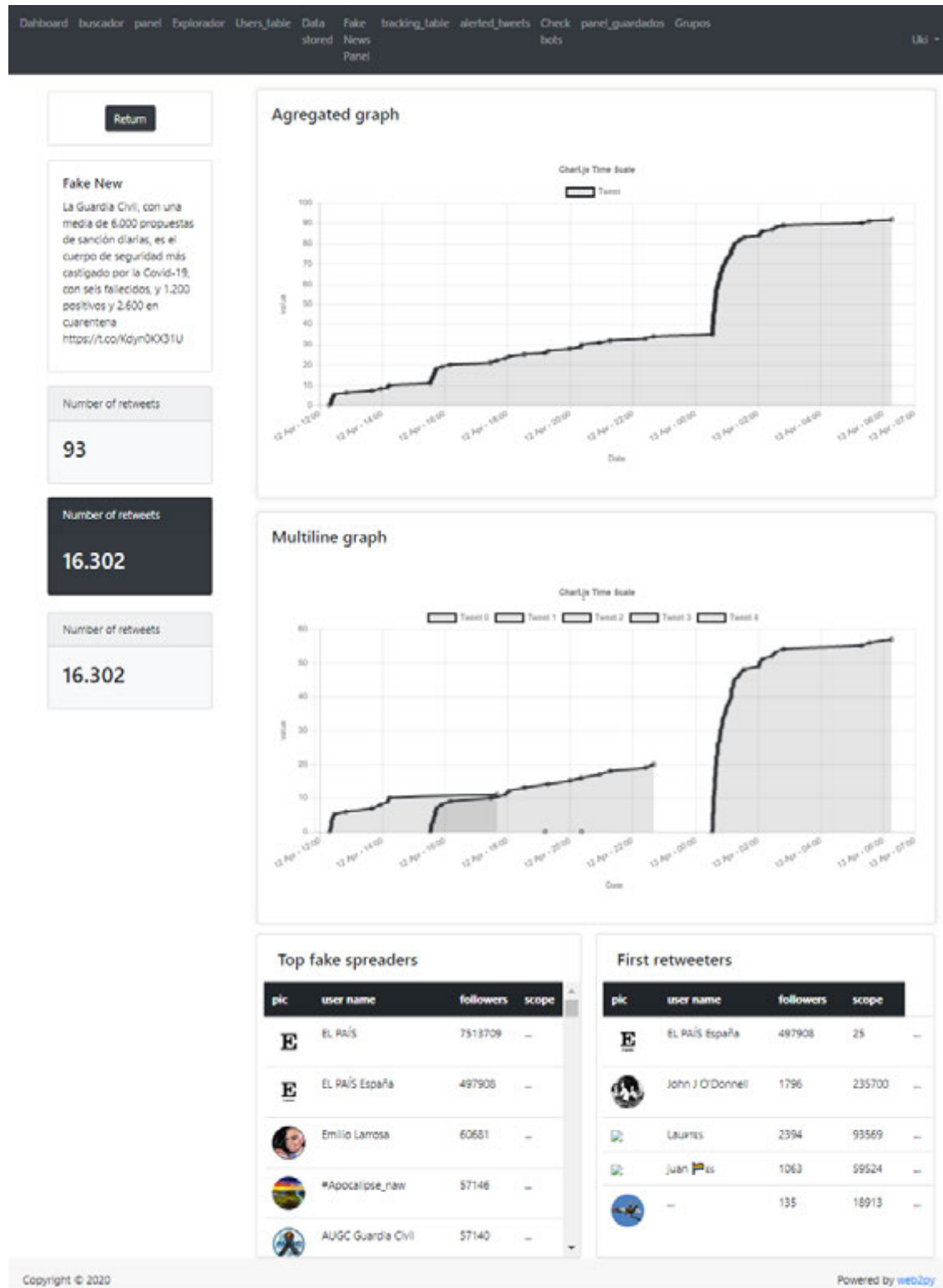


Ilustración 48. Visualización de tweets agregados

Base de datos, agrupación de Tweets

De igual forma que se pueden buscar Tweets individuales, también puede visualizarse en la base de datos aquellas agrupaciones de Tweets, que se componen del id principal (Que se corresponde con el id del tweet base sobre el que se buscaron los demás), de los ids de los tweets similares, y otra serie de opciones como la visualización de si el grupo está en modo tracking activado, si tiene actualizaciones no vistas, u opciones de ver editar y borrar. Además, cuenta con la opción de ir al panel de análisis, donde se puede examinar cada grupo de la forma mencionada anteriormente.

Search term	Search user	Relevant accounts	Stored tweets	Grouped tweets	Check bots	Fake News Panel	Ukl
-------------	-------------	-------------------	---------------	----------------	------------	-----------------	-----

Grouped Tweets

5 records found

Main Id	Ids	Tracking	Has Updates	User	Text	Analysis panel
92	95, 96, 98	<input type="checkbox"/>	<input type="checkbox"/>	abc_es	#LoMásLeído 📰 Hacienda ultima aplazar impuestos a pymes y autónomos que facturen hasta 600.000 euros https://t.co/aXDZP7EhDq	<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>
102	105, 106, 107, 13...	<input type="checkbox"/>	<input type="checkbox"/>	elmundoes	💎 Casi 160.000 casos confirmados del nuevo #coronavirus en el estado de Nueva York, más que en cualquier país del mundo, y cerca de 8.000 muertos. ¿Por qué este estado de 19 millones de habitantes ha sido tan afectado por la #pandemia? 📰 https://t.co/wU9YcFtRAq	<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>
114	116, 117	<input type="checkbox"/>	<input type="checkbox"/>	elpais_espaa	El presidente reconoce en el Congreso errores pero pide unidad y lealtad: "Toda Europa llegó tarde, pero España actuó antes" https://t.co/de7q8h5NqD	<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>
218	220, 222, 372, 373	<input type="checkbox"/>	<input type="checkbox"/>	elpais_espaa	La Guardia Civil, con una media de 6.000 propuestas de sanción diarias, es el cuerpo de seguridad más castigado por la Covid-19, con seis fallecidos, y 1.200 positivos y 2.600 en cuarentena https://t.co/Kdyn0KX31U	<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>
112	None	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	okdiario	📰 @EduardoIñda: «Los Pactos de la Moncloa no pueden servir para mantener al culpable de este desaguado en La Moncloa» https://t.co/0EUpD2D868	<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>

Ilustración 49. Base de datos, consulta de tweets agregados

Análisis de cuentas bots

Por último, el panel de análisis de cuentas bot, que permite introducir una cuenta de Twitter, y devuelve una serie de resultados, como la probabilidad de que una cuenta sea bot, y una puntuación al nivel de actividad y al nivel de análisis de sentimientos de la cuenta.

Search term Search user Relevant accounts Stored tweets Grouped tweets Check bots Fake News Panel User

Check bots

Introduce an account for checking the probability of being a bot

Check Bot

Sentiment score (%)	Bot probability (%)	Activity score (%)
48	38	93

Copyright © 2020 Powered by web2py

Ilustración 50. Pantalla de control de bots

8.2. Futuras vías de exploración

Un famoso dicho dentro de la comunidad de desarrollo de software menciona que “un proyecto de software nunca termina”. En mi opinión, es una frase muy acertada, puesto que, aunque se defina el alcance de un proyecto de software de forma clara y anticipada, a medida que avanza el desarrollo y tras el mismo van surgiendo nuevas ideas y mejoras que pueden enriquecer el sistema, (al margen de la necesidad de mantenimiento de un sistema y de su actualización a medida que la tecnología avanza).

Este proyecto no es una excepción, por lo que a continuación se exponen las vías en las que se quiere seguir trabajando, pero quedan fuera del alcance inicial planteado.

1. **Uso de subsistemas de inteligencia artificial para la detección de fake news.** Para la primera versión del proyecto no se han construido este tipo de sistemas de detección automática de fake news. El motivo ha sido que la construcción de estos algoritmos o sistemas son complejos en sí mismos y pueden abarcar el alcance de un proyecto completo o incluso

de un doctorado. Sin embargo, sí existe la posibilidad de conectar el sistema en el futuro con otros proyectos más avanzados en este sentido, o crear algoritmos de detección propios e implementarlos en una segunda versión de la aplicación.

2. **Extrapolar el sistema a otras redes sociales.** El sistema construido se ha basado en Twitter, por ser una de las redes sociales que más facilidad ofrecen a la comunidad de desarrolladores para explotar su API y extraer información, pero no es la única red social que lo permite. En el futuro se desea incluir la posibilidad de conectar otras redes sociales en el sistema y ampliar el alcance de la herramienta.
3. **Mejora de las interfaces y funcionalidad descrita:** Existen bastantes posibilidades de mejora respecto al sistema actual, como la creación de nuevos diseños para las interfaces o la mejora de la experiencia de usuario.

8.3. Limitaciones del proyecto

Junto con los resultados obtenidos, considero necesario reflejar las limitaciones que tiene el proyecto que se presenta, que se centran en tres aspectos principales.

En primer lugar, el sistema está limitado por el uso de la **versión básica de la API de Twitter**: El sistema tiene como base en sí mismo la necesidad de conseguir información para que pueda ser mostrada y analizada. En este caso la aplicación usa la versión básica que Twitter ofrece a través de su API. Como principal ventaja es que se trata de un servicio gratuito, y que cubre las necesidades suficientes como para conseguir una aplicación web funcional y demostrar resultados. Como aspecto negativo, es que la versión gratuita tiene algunas limitaciones, como limitaciones en el número de tweets obtenido, o límites en las conexiones que se pueden establecer con Twitter vía API.

En segundo lugar, la **dependencia tecnológica**. El sistema está basado en el uso de otras tecnologías y APIs que están en continua evolución, por lo que una vez construido el sistema es necesario contar con un mantenimiento adecuado para asegurar que siga funcionando de forma correcta.

Y en tercer lugar, pero en mi opinión la limitación más importante: **Limitaciones legales**. El sistema tiene como pilar principal la capacidad de extracción de información de Twitter, pero esto trae consigo la necesidad de prestar atención a las limitaciones legales sobre la obtención y uso de dicha información, debiendo ser conocedores en todo momento de los términos y condiciones de Twitter, así

como de la regulación en relación a la privacidad de los usuarios. Este punto es especialmente relevante porque limita el alcance y capacidades del sistema, y porque puede acarrear consecuencias legales.

8.4. Impacto socio económico

Se hace difícil hacer un estudio concreto del impacto socio económico que el proyecto puede tener. En la primera versión del sistema que se presenta, no parece viable una explotación que permita cubrir los costes acarreados de la construcción del mismo reflejados en secciones anteriores. Pero sí que sirve como base para futuros sistemas, y en segundas y terceras versiones se abren algunas ventanas de oportunidad para la explotación del mismo.

Desde el punto de vista económico, las principales vías de explotación del proyecto son tres:

- **Explotación del software como servicio.** El sistema podría ser monetizado mediante el pago de cuotas mensuales por parte de los usuarios para su explotación. Para ello es necesario prestar atención a la capacidad de escalabilidad del sistema, y a crear una aplicación lo más amplia posible para abrir el abanico de posibles clientes que puedan cubrir los costes reflejados en el apartado de gestión del proyecto. Esto podría conseguirse mediante un modelo freemium, que permita una rápida expansión del mismo, ofreciendo servicios gratuitos por defectos, y una funcionalidad completa a aquellos usuarios que estén interesados en la versión más avanzada del sistema.
- **Como servicio de información.** En mi opinión, la opción más viable, que consiste en ofrecer un servicio de información y análisis para terceros en un modelo de consultoría, en el que se usa la herramienta como medio de análisis, y ofrecer los resultados obtenidos como principal valor para los clientes.
- Por último, parece interesante investigar **subvenciones y programas que ofrecen los gobiernos** para hacer frente al problema de las noticias falsas. Actualmente los gobiernos de Europa y del resto del mundo están creando planes de acción [35] para la lucha contra la desinformación y cuentan con varios programas dedicados a promover soluciones a este problema, y ofrecen incentivos para ello a través de subvenciones.

9. CONCLUSIONES

En este apartado se reflejan las principales conclusiones obtenidas no sólo respecto al proyecto en si mismo, sino también con respecto a la experiencia y proceso de realización del mismo.

En primer lugar, reflejar que los resultados del proyecto han sido (en mi opinión) muy positivos. El objetivo de escoger hacer una aplicación web para el proyecto de fin de carrera era aprender varias de las tecnologías que se han ido usando (Vue.js, Web2py, uso de APIs, etc) y construir un producto final que sirviese como prueba de concepto o mínimo producto viable para explorar las posibilidades de la idea. En este sentido la aplicación construida ha cubierto los objetivos mencionados.

En segundo lugar, si tuviera que repetir el proceso desde el principio, creo que seguiría un enfoque algo diferente. Cuando me planteé el proyecto, mi idea principal era aprovechar un proyecto de envergadura para aprender en algo que estuviera interesado. Debido a esto he puesto una parte muy importante del tiempo en la construcción de la aplicación en si misma por delante del documento que se expone. Creo que esto ha hecho que en ocasiones pierda el foco de que se trataba de un trabajo principalmente académico, en el que probablemente se centra más el foco en la documentación que en la construcción en sí misma. El tiempo invertido en todo lo relacionado con el proyecto durante este año ha bastante elevado, poner el foco en obtener una aplicación real ha hecho que este año haya tenido que gestionar el tiempo de forma muy eficiente entre mi trabajo exigente en el mundo laboral (Accenture) y el trabajo académico. No obstante, pienso que el resultado final ha sido muy positivo, y que he conseguido mi objetivo personal de aprender y construir un producto funcional, además del objetivo académico de documentar el proyecto de la forma adecuada.

En tercer lugar, pienso que, en caso de repetir el proyecto, escogería hacerlo con otra serie de tecnologías totalmente diferentes. No sólo porque alguna de ellas como Web2py es una tecnología que se queda algo antigua, sino porque considero que en esta etapa es imprescindible conocer un espectro amplio de tecnologías, y ser capaz de aprender de una forma lo más eficiente posible. Al fin y al cabo, en el mundo del desarrollo de software existen tantas tecnologías y avanzan de forma tan rápida, que al final de las principales cosas que se valoran de un profesional del sector son su la actitud, la capacidad de adaptación y la capacidad de aprender.

Por último, mencionar que este proyecto comenzó a mediados de octubre de 2019, sin conocer la situación excepcional que venía por delante. La pandemia mundial ha reflejado aún más la importancia de la lucha contra las noticias falsas. A lo largo del estado de alarma (especialmente al comienzo del

mismo) el aumento del uso de las redes sociales ha sido notable, y con ello el aumento de las noticias falsas y bulos distribuidos en las redes. Tanto es así que las propias redes sociales han tomado ciertas medidas para evitar la desinformación, y el tema de las noticias falsas y el cómo hacerles frente ha sido un tema que se ha elevado incluso al nivel político, creando en ocasiones algunas polémicas como se ha reflejado en las primeras secciones del documento.

10. BIBLIOGRAFÍA

- [1] K. E. Matsa. E. Shearer, "News Use Across Social Media Platforms 2018," *Pew Research Center*, 2018 [En línea]. Disponible en: <https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/>.
- [2] V. A. Martinez-Fernandez, O. Juanatey-Boga, V. Crespo-Pereira y M. D. Mahauad Burneo, "Prensa nativa digital y redes sociales" *International Journal of Information Systems and Software Engineering for Big Companies (IJISEBC)*, 2015.
- [3] M. Hindman, "Knight Foundation," october 2018. [En línea]. Disponible en: https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/238/original/KF-DisinformationReport-final2.pdf. [Último acceso: 25 june 2020].
- [4] S. Aral, D. Roy, S. Vosoughi, "The spread of true and false news online," *Science*, vol. 359, nº 6380, p. 7, 9 march 2018.
- [5] J. Clement, "statista.com," 20 febrero, 2020. [En línea]. Disponible en: <https://www.statista.com/topics/737/twitter/>. [Último acceso: 20 junio 2020].
- [6] DE-CIX Madrid statistics, "DE-CIX Madrid statistics," june 2020. [En línea]. Disponible en: <https://www.de-cix.net/en/locations/spain/madrid/statistics>. [Último acceso: junio 2020].
- [7] S. Perez, "techcrunch," 4 june, 2020. [En línea]. Disponible en: <https://techcrunch.com/2020/06/04/twitter-has-a-record-breaking-week-as-users-looked-for-news-of-protests-and-covid-19/?guccounter=1>. [Último acceso: 25 junio 2020].
- [8] I. Zulet, "La Guardia Civil tiene la misión de "minimizar las críticas" a Sánchez en la red, según su 'número dos,'" *El Español*, 19 abril 2020.
- [9] P. Ortega y M. González, "Tormenta política tras declarar un general que la Guardia Civil trabaja para minimizar bulos contrarios al Gobierno," *El Pais*, 2020.

- [10] EFE Madrid, “Marlaska afea a PP y Vox que usen el lapsus del general de la Guardia Civil para criticar a Interior,” *La Vanguardia*, 20 abril, 2020.
- [11] Fondo Monetario Internacional, “Informes de Perspectivas de Economía Mundial, abril de 2020” *FMI*, abril 2020. [En línea]. Disponible en: <https://www.imf.org/es/Publications/WEO/Issues/2020/04/14/weo-april-2020>. [Último acceso: 25 junio 2020].
- [12] A. Robinson, “El PIB de España caerá un 8% este año y el paro rozará el 21%, según el FMI,” *La Vanguardia*, 14 abril, 2020.
- [13] EFE, “El FMI dice que las vulnerabilidades de España acentúan la crisis económica,” *Expansion*, 15 abril, 2020.
- [14] B. Marr, “How The COVID-19 Pandemic Is Fast-Tracking Digital Transformation In Companies,” *Forbes*, 2020.
- [15] E. Schulze, “EU tells Facebook, Google and Twitter to take more action on fake news,” *CNBC*, 29 october, 2019.
- [16] European Union, “General data protection regulation, Regulation (EU) 2016/679,” 2016.
- [17] California US State, “California Consumer Privacy Act (CCPA)” 2020.
- [18] Deloitte Consultoría de Riesgos , “Deloitte España,” [En línea]. Disponible en: [https://www2.deloitte.com/content/dam/Deloitte/co/Documents/risk/Brochure_Entendiendo%20el%20Reglamento%20General%20de%20Protecci%C3%B3n%20de%20Datos%20\(GDPR\)_A4_Col.pdf](https://www2.deloitte.com/content/dam/Deloitte/co/Documents/risk/Brochure_Entendiendo%20el%20Reglamento%20General%20de%20Protecci%C3%B3n%20de%20Datos%20(GDPR)_A4_Col.pdf). [Último acceso: may 2020].
- [19] IT User, “8 claves para entender la GDPR: entramos en la recta final,” *ituser*, 11 mayo, 2018. [En línea]. Disponible en: <https://www.ituser.es/seguridad/2018/05/8-claves-para-entender-gdpr-entramos-en-la-recta-final>. [Último acceso: mayo 2020].
- [20] “Twitter privacy policy,” Twitter. [En línea]. Disponible en: <https://twitter.com/en/privacy>. [Último acceso: may 2020].

- [21] "Twitter developer terms and policy," Twitter, [En línea]. Disponible en: <https://developer.twitter.com/en/developer-terms/policy.html>. [Último acceso: may 2020].
- [22] "Twitter automation rules," Twitter, [En línea]. Disponible en: <https://help.twitter.com/en/rules-and-policies/twitter-automation>. [Último acceso: 26 may 2020].
- [23] A. Mosseri, "Working to stop misinformation and false news," *Facebook*, 7 abril 2017. [En línea]. Disponible en: <https://www.facebook.com/facebookmedia/blog/working-to-stop-misinformation-and-false-news>. [Último acceso: mayo 2020].
- [24] BBC, "Whatsapp restringes message-sharing to fight fake news," *BBC*, 21 january 2019. [En línea]. Disponible en: <https://www.bbc.com/news/technology-46945642>. [Último acceso: may 2020].
- [25] P. Agrawal, "Twitter acquires Fabula AI to strengthen its machine learning expertise," *Blog oficial de Twitter*, 3 de Junio, 2019.
- [26] ht Correspondent, "Twitter confirms it is working on these features to fight spreading of misinformation," 21 febrero, 2020. [En línea]. Disponible en: <https://tech.hindustantimes.com/tech/news/twitter-confirms-it-is-working-on-these-features-to-fight-spreading-of-misinformation-story-uHhcMttunfGwsIRsm4mD2N.html> [Último acceso: 25 mayo 2020].
- [27] A. Bondielli y F. Marcellon, "A survey on fake news and rumour detection techniques," *Information Sciences*, pp. 38-55, 2019.
- [28] K. We, S. Yang y K. Q., "False rumors detection on Sina Weibo by propagation structures," de *IEEE 31st international conference on data engineering*, 2015.
- [29] S. Hamidian y M. T. Diab, "Rumor Detection and Classification for Twitter Data," *arXiv preprint arXiv:1912.08926*, 2019.
- [30] S. Kwon, M. Cha, K. Jung, W. Chen y Y. Wang, "'Prominent Features of Rumor Propagation in Online Social Media'," de *IEEE 13th International Conference on Data Mining, Dallas*, 2013.

- [31] M. Hardalov, I. Koychev y P. Nakov, "Hardalov, Momchil, Ivan Koychev, and Preslav Nakov," de *International Conference on Artificial Intelligence: Methodology, Systems, and Applications.*, Cham, 2016.
- [32] Y.-C. Chen, L. Zhao-Yang y K. Hung-Yu, "IKM at SemEval-2017 Task 8: Convolutional Neural Networks for," de *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Tainan, Taiwan, 2017.
- [33] N. Maja, "Reasons Why Vue.js Is Getting More Traction Every Month," *Monterail blog*, 2018.
- [34] M. Di Pierro, "Why web2py," de *Web2py, Complete Reference Manual*, 2013.
- [35] European Comission. (EC), "Action Plan against Disinformation.," 2018.
- [36] J. Noguera, "Redes sociales como paradigma periodístico. Medios españoles en Facebook," *Revista Latina de comunicacion*, 2010.

11. ANEXO

11.1. DECLARACIÓN DE ORIGINALIDAD

DECLARACION DE ORIGINALIDAD

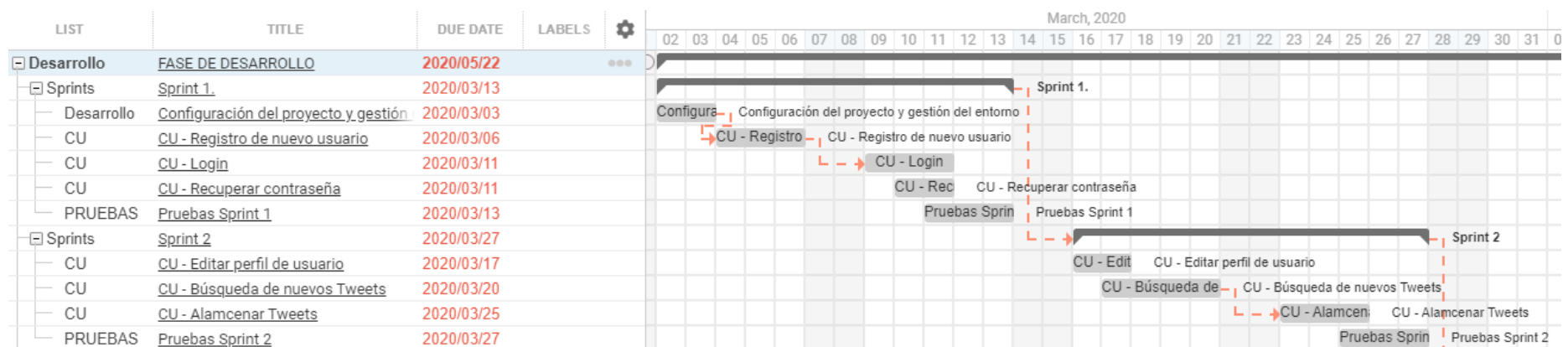
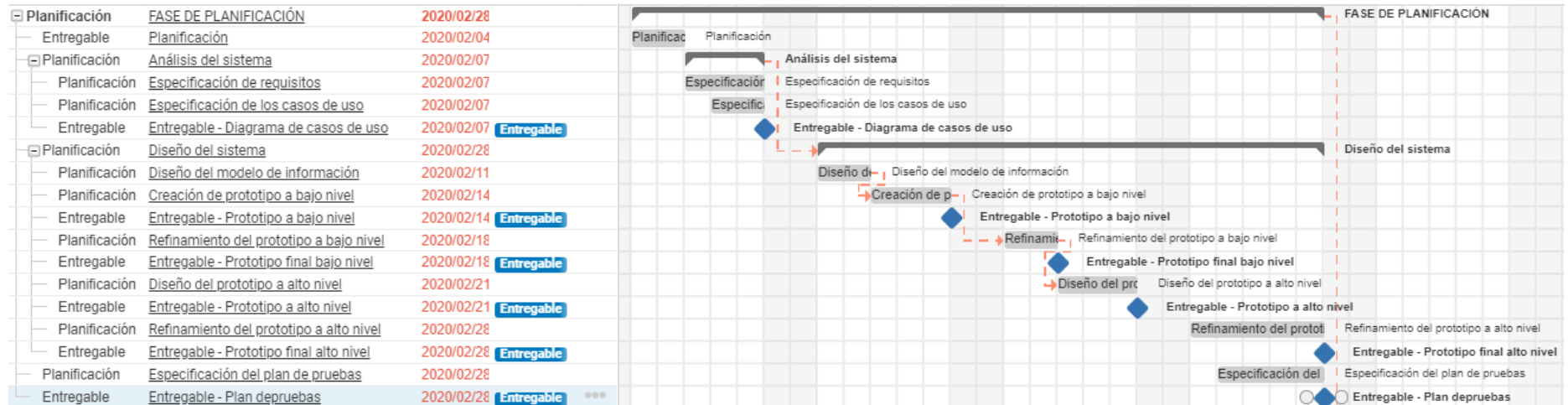
Yo, ...Ignacio Laviña Faustmann... declaro que el TFG "**Herramienta para el análisis de difusión de noticias falsas en Twitter**" es totalmente original mío, que no ha sido presentado en ninguna otra universidad como TFG y que todas las fuentes que han sido utilizadas han sido adecuadamente citadas y aparecen en las referencias bibliográficas.

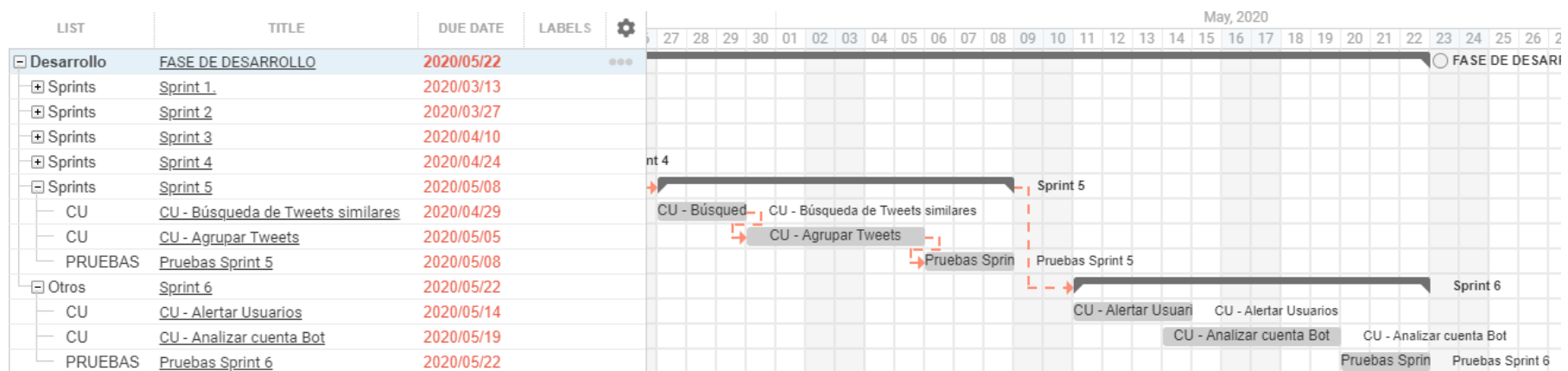
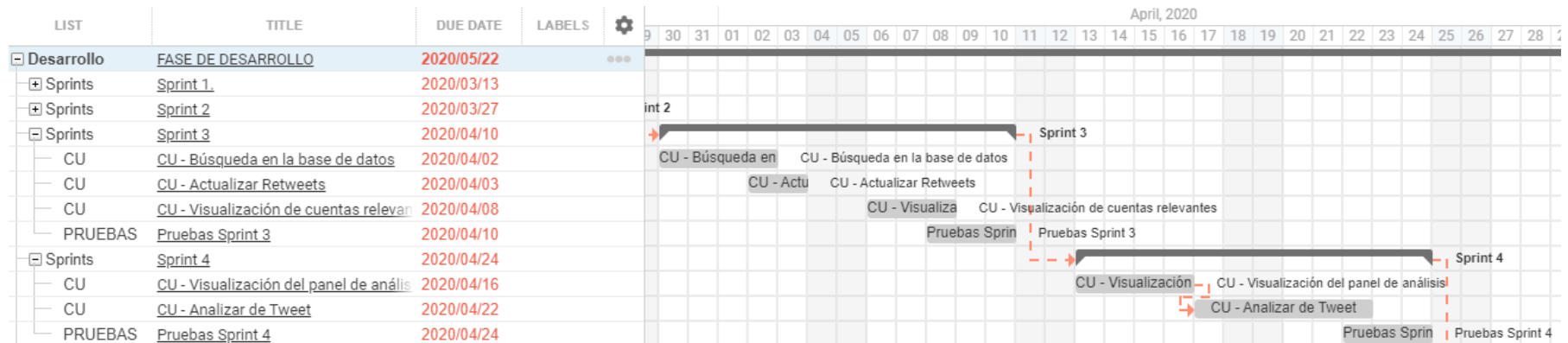
Getafe/Colmenarejo, a 5 de Julio de 2020

Firma:

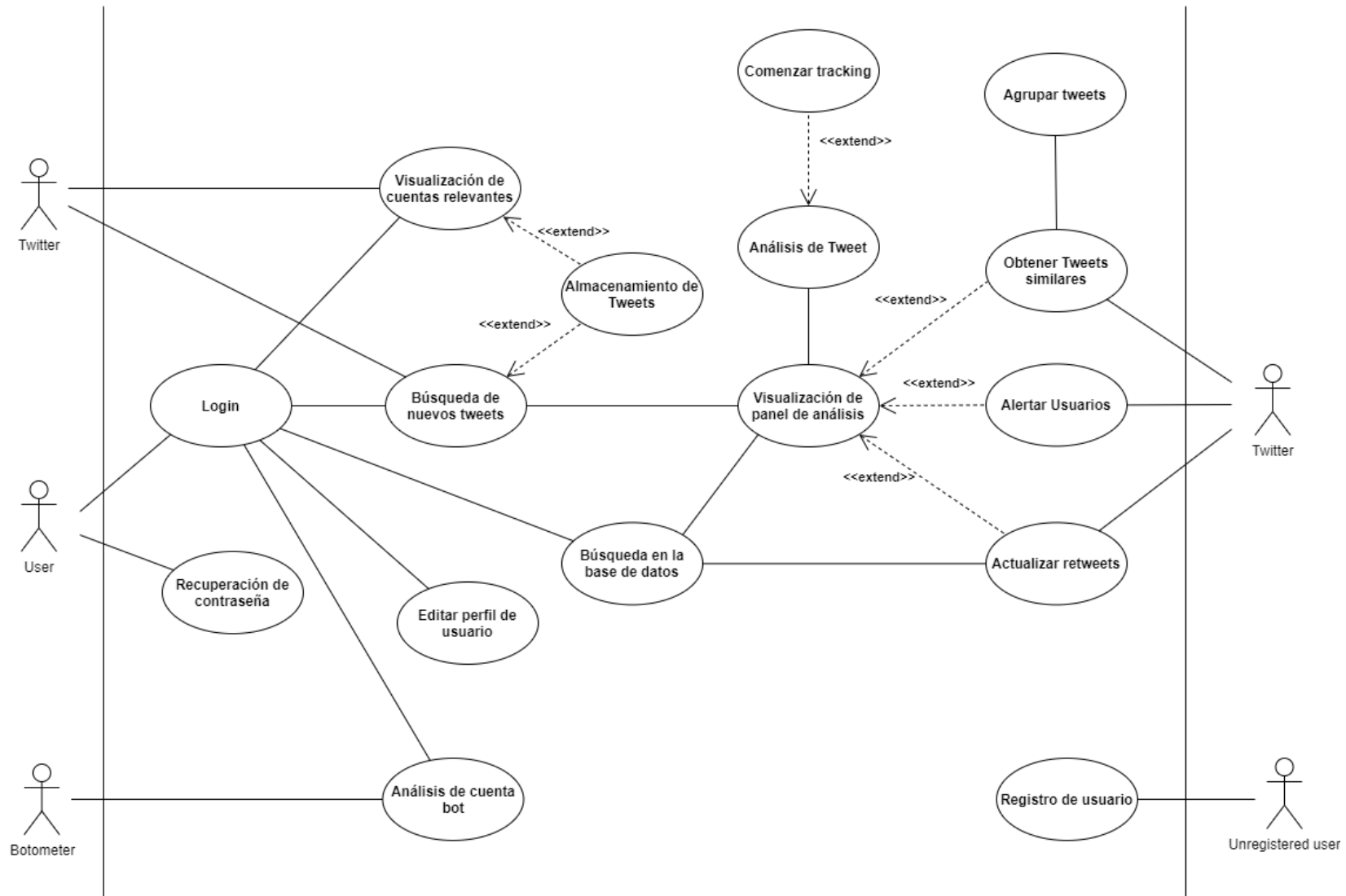


11.2. DIAGRAMA DE GANTT

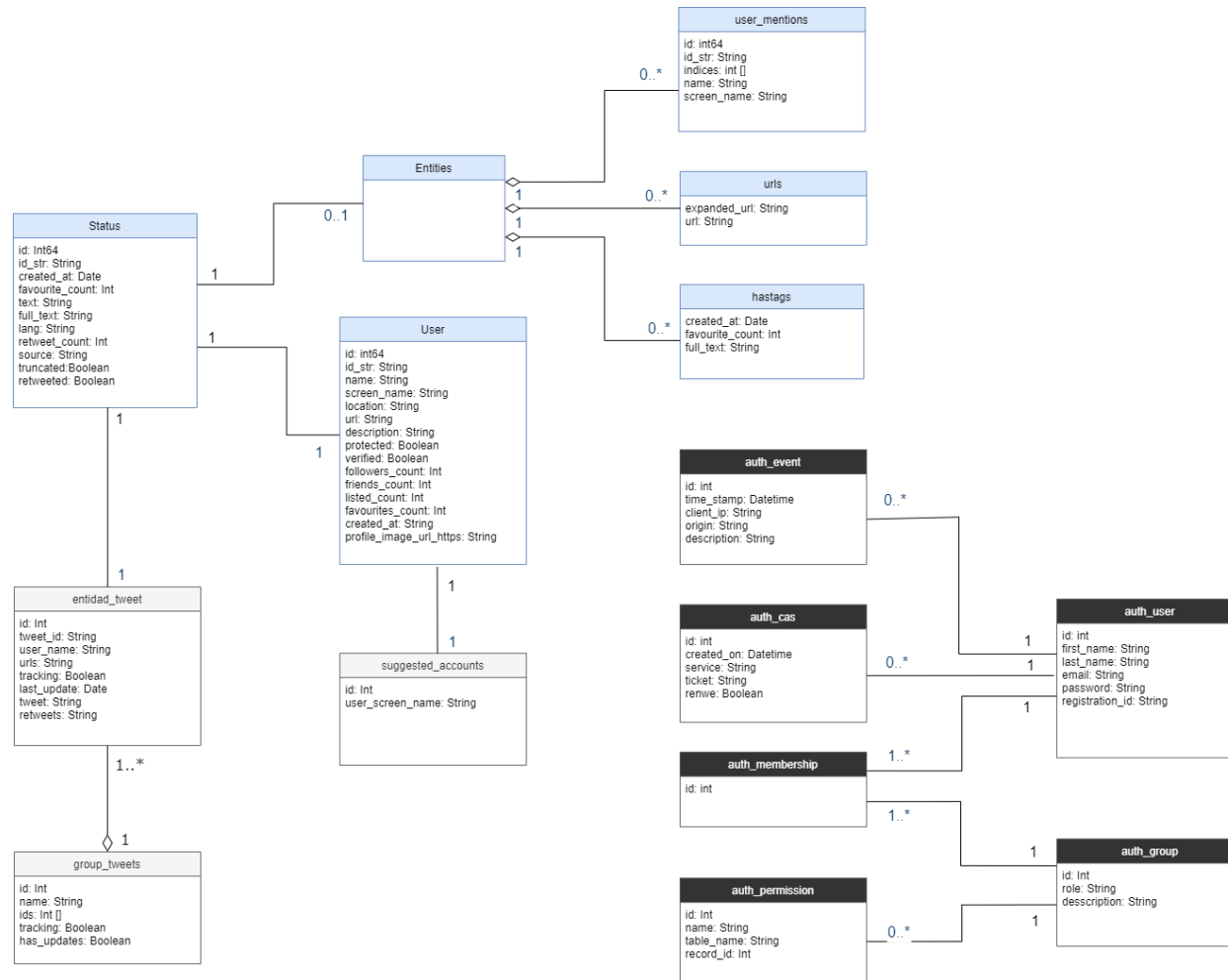




11.3. DIAGRAMA DE CASOS DE USO



11.4. MODELO DE INFORMACIÓN



11.4.1. DIAGRAMA DE FLUJO PROTOTIPO

